GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 12:45:42 ; Search time 15045.6 Seconds
(without alignments)
5398.140 Million cell updates/sec

Title: US-09-410-835-4
Perfect score: 5925
Sequence: 1 ctccattttaagcaactcca.....ttttagaatagacaaagctt 5925

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1118133 seqs, 6853842396 residues

Word size: 15

Total number of hits satisfying chosen parameters: 1118785

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*

22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vi:*
29: gb_htg1:*
30: gb_htg2:*
31: gb_in1:*
32: gb_in2:*
33: em_ba1:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: em_htg1:*
44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_htg8:*
50: gb_htg9:*
51: gb_htg10:*
52: gb_htg11:*
53: gb_htg12:*
54: gb_htg13:*
55: gb_htg14:*
56: gb_in3:*
57: gb_htg15:*
58: gb_htg16:*
59: gb_htg17:*
60: em_htg4:*
61: em_htg5:*
62: em_htg6:*
63: em_htg7:*
64: em_hum6:*
65: gb_htg18:*
66: gb_htg19:*
67: gb_htg20:*
68: gb_htg21:*
69: gb_htg22:*
70: gb_htg23:*
71: gb_vi1:*
72: gb_vi2:*

73: gb_ba3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

length of match

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5925	100.0	5925	2	AX030324	AX030324 Sequence
2	5925	100.0	5925	2	HPCAI	X70039 H.pylori ca
3	5925	100.0	5925	81	A76409	A76409 Sequence 4
4	5642	95.2	20229	2	AF282853	AF282853 Helicobac
5	353	6.0	413	2	HPY239688	AJ239688 Helicobac
6	316	5.3	413	2	HPY239685	AJ239685 Helicobac
7	289	4.9	413	2	HPY239686	AJ239686 Helicobac
8	288	4.9	449	2	HPY269887	AJ269887 Helicobac
9	277	4.7	449	2	HPY269894	AJ269894 Helicobac
10	277	4.7	449	2	HPY269895	AJ269895 Helicobac
11	272	4.6	449	2	HPY269881	AJ269881 Helicobac
12	259	4.4	413	2	HPY239683	AJ239683 Helicobac
13	243	4.1	243	2	HPY239721	AJ239721 Helicobac
14	243	4.1	243	2	HPY239730	AJ239730 Helicobac
15	235	4.0	291	2	AF289431	AF289431 Helicobac
16	235	4.0	3648	82	I95673	I95673 Sequence 1
17	235	4.0	4821	82	I95674	I95674 Sequence 3
18	235	4.0	4826	2	HECMAJANT	L11714 Helicobacte
19	228	3.8	3540	1	AF247651	AF247651 Helicobac
20	222	3.7	1100	1	AF191014	AF191014 Helicobac
21	221	3.7	413	2	HPY239687	AJ239687 Helicobac
22	221	3.7	413	2	HPY239689	AJ239689 Helicobac
23	212	3.6	576	2	AF289433	AF289433 Helicobac
24	210	3.5	449	2	HPY269896	AJ269896 Helicobac
25	194	3.3	449	2	HPY269891	AJ269891 Helicobac
26	185	3.1	1233	1	AF190659	AF190659 Helicobac
27	182	3.1	11141	1	AE000569	AE000569 Helicobac
28	181	3.1	787	1	AF190663	AF190663 Helicobac

29	179	3.0	449	2	HPY269879	AJ269879 Helicobac
30	179	3.0	449	2	HPY269882	AJ269882 Helicobac
31	175	3.0	413	2	HPY239691	AJ239691 Helicobac
32	175	3.0	449	2	HPY269888	AJ269888 Helicobac
33	174	2.9	449	2	HPY269883	AJ269883 Helicobac
34	169	2.9	287	1	AF202220	AF202220 Helicobac
35	167	2.8	672	2	AF289438	AF289438 Helicobac
36	166	2.8	10781	1	AE001483	AE001483 Helicobac
37	158	2.7	291	2	AF289406	AF289406 Helicobac
38	158	2.7	291	2	AF289417	AF289417 Helicobac
39	158	2.7	4093	1	AF083352	AF083352 Helicobac
40	157	2.6	291	2	AF289422	AF289422 Helicobac
41	157	2.6	413	2	HPY239684	AJ239684 Helicobac
42	156	2.6	449	2	HPY269889	AJ269889 Helicobac
43	154	2.6	244	2	HPY252970	AJ252970 Helicobac
44	154	2.6	286	1	AF198472	AF198472 Helicobac
45	153	2.6	220	1	AF198479	AF198479 Helicobac
46	152	2.6	1323	1	AF222809	AF222809 Helicobac
47	149	2.5	413	2	HPY239692	AJ239692 Helicobac
48	144	2.4	291	2	AF289420	AF289420 Helicobac
49	141	2.4	291	2	AF289400	AF289400 Helicobac
50	140	2.4	220	1	AF198482	AF198482 Helicobac
51	140	2.4	243	2	HPY239736	AJ239736 Helicobac
52	140	2.4	281	1	AF198468	AF198468 Helicobac
53	140	2.4	290	1	AF198477	AF198477 Helicobac
54	140	2.4	413	2	HPY239690	AJ239690 Helicobac
55	138	2.3	449	2	HPY269884	AJ269884 Helicobac
56	138	2.3	3744	1	AB003397	AB003397 Helicobac
57	138	2.3	3744	1	AF001357	AF001357 Helicobac
58	138	2.3	3744	1	AF202973	AF202973 Helicobac
59	137	2.3	220	1	AF198474	AF198474 Helicobac
60	137	2.3	243	2	HPY239732	AJ239732 Helicobac
61	137	2.3	243	2	HPY239737	AJ239737 Helicobac
62	137	2.3	244	2	HPY252968	AJ252968 Helicobac
63	137	2.3	244	2	HPY252969	AJ252969 Helicobac
64	137	2.3	244	2	HPY252971	AJ252971 Helicobac
65	137	2.3	244	2	HPY252974	AJ252974 Helicobac
66	137	2.3	281	1	AF198469	AF198469 Helicobac
67	137	2.3	287	1	AF202221	AF202221 Helicobac
68	137	2.3	299	1	AF198471	AF198471 Helicobac
69	136	2.3	449	2	HPY269880	AJ269880 Helicobac
70	132	2.2	291	2	AF289402	AF289402 Helicobac
71	131	2.2	449	2	HPY269886	AJ269886 Helicobac
72	130	2.2	413	2	HPY239693	AJ239693 Helicobac
73	126	2.1	413	2	HPY239694	AJ239694 Helicobac
74	125	2.1	449	2	HPY269890	AJ269890 Helicobac
75	124	2.1	290	1	AF198475	AF198475 Helicobac
76	123	2.1	291	2	AF289411	AF289411 Helicobac
77	121	2.0	449	2	HPY269885	AJ269885 Helicobac
78	119	2.0	243	2	HPY239729	AJ239729 Helicobac
79	119	2.0	449	2	HPY269897	AJ269897 Helicobac

	80	118	2.0	291	2	AF289407	AF289407 Helicobac
	81	118	2.0	678	2	AF289456	AF289456 Helicobac
	82	116	2.0	540	2	AF289437	AF289437 Helicobac
	83	116	2.0	567	2	AF289450	AF289450 Helicobac
	84	116	2.0	567	2	AF289457	AF289457 Helicobac
	85	115	1.9	284	1	AF198476	AF198476 Helicobac
	86	114	1.9	2963	1	AF190660	AF190660 Helicobac
	87	111	1.9	291	2	AF289399	AF289399 Helicobac
	88	111	1.9	291	2	AF289416	AF289416 Helicobac
	89	110	1.9	291	2	AF289418	AF289418 Helicobac
	90	109	1.8	291	2	AF289429	AF289429 Helicobac
	91	109	1.8	987	1	AF190658	AF190658 Helicobac
	92	108	1.8	464	2	HPY269875	AJ269875 Helicobac
	93	108	1.8	1335	1	AF222807	AF222807 Helicobac
	94	107	1.8	291	2	AF289415	AF289415 Helicobac
	95	107	1.8	291	2	AF289430	AF289430 Helicobac
	96	107	1.8	590	1	AF190996	AF190996 Helicobac
	97	104	1.8	956	1	AF190992	AF190992 Helicobac
	98	101	1.7	567	2	AF289442	AF289442 Helicobac
	99	101	1.7	678	2	AF289435	AF289435 Helicobac
	100	101	1.7	1224	1	AF222808	AF222808 Helicobac
	101	100	1.7	291	2	AF289403	AF289403 Helicobac
	102	100	1.7	291	2	AF289405	AF289405 Helicobac
	103	100	1.7	291	2	AF289409	AF289409 Helicobac
	104	100	1.7	291	2	AF289410	AF289410 Helicobac
	105	100	1.7	291	2	AF289413	AF289413 Helicobac
	106	100	1.7	291	2	AF289419	AF289419 Helicobac
	107	100	1.7	291	2	AF289421	AF289421 Helicobac
	108	100	1.7	291	2	AF289425	AF289425 Helicobac
	109	100	1.7	291	2	AF289427	AF289427 Helicobac
	110	100	1.7	291	2	AF289428	AF289428 Helicobac
	111	98	1.7	291	2	AF289408	AF289408 Helicobac
	112	97	1.6	897	1	AF190995	AF190995 Helicobac
	113	96	1.6	449	2	HPY269892	AJ269892 Helicobac
	114	96	1.6	449	2	HPY269893	AJ269893 Helicobac
	115	96	1.6	1000	1	AF190993	AF190993 Helicobac
	116	94	1.6	940	1	AF190994	AF190994 Helicobac
	117	94	1.6	3522	1	AF202972	AF202972 Helicobac
	118	92	1.6	243	2	HPY239723	AJ239723 Helicobac
	119	89	1.5	286	1	AF202224	AF202224 Helicobac
	120	89	1.5	291	2	AF289401	AF289401 Helicobac
	121	89	1.5	291	2	AF289412	AF289412 Helicobac
	122	89	1.5	291	2	AF289414	AF289414 Helicobac
	123	89	1.5	291	2	AF289423	AF289423 Helicobac
	124	88	1.5	291	2	AF289426	AF289426 Helicobac
	125	88	1.5	678	2	AF289441	AF289441 Helicobac
c	126	88	1.5	10351	1	AE000560	AE000560 Helicobac
	127	87	1.5	291	2	AF289424	AF289424 Helicobac
	128	86	1.5	243	2	HPY239734	AJ239734 Helicobac
	129	86	1.5	243	2	HPY239735	AJ239735 Helicobac
	130	86	1.5	244	2	HPY252963	AJ252963 Helicobac

	131	86	1.5	244	2	HPY252964	AJ252964 Helicobac
	132	86	1.5	286	1	AF202223	AF202223 Helicobac
	133	86	1.5	784	1	AF191013	AF191013 Helicobac
c	134	86	1.5	2693	1	AF125214	AF125214 Helicobac
	135	85	1.4	567	2	AF289458	AF289458 Helicobac
	136	85	1.4	570	2	AF289439	AF289439 Helicobac
	137	85	1.4	576	2	AF289447	AF289447 Helicobac
	138	85	1.4	576	2	AF289461	AF289461 Helicobac
	139	85	1.4	576	2	AF289463	AF289463 Helicobac
	140	85	1.4	633	2	AF289453	AF289453 Helicobac
	141	85	1.4	633	2	AF289455	AF289455 Helicobac
	142	85	1.4	672	2	AF289434	AF289434 Helicobac
	143	85	1.4	672	2	AF289451	AF289451 Helicobac
	144	84	1.4	570	2	AF289443	AF289443 Helicobac
	145	84	1.4	573	2	AF289462	AF289462 Helicobac
	146	84	1.4	676	2	AF289449	AF289449 Helicobac
	147	84	1.4	678	2	AF289432	AF289432 Helicobac
	148	83	1.4	669	2	AF289446	AF289446 Helicobac
	149	80	1.4	220	1	AF198481	AF198481 Helicobac
	150	80	1.4	294	1	AF198470	AF198470 Helicobac
	151	78	1.3	567	2	AF289460	AF289460 Helicobac
	152	77	1.3	2378	2	HPU95957	U95957 Helicobacte
	153	76	1.3	291	2	AF289404	AF289404 Helicobac
	154	76	1.3	570	2	AF289436	AF289436 Helicobac
	155	73	1.2	269	2	HPU80065	U80065 Helicobacte
	156	73	1.2	452	2	HPY269866	AJ269866 Helicobac
c	157	73	1.2	590	1	AF190996	AF190996 Helicobac
	158	73	1.2	11680	1	AE000609	AE000609 Helicobac
	159	73	1.2	16693	1	AE000557	AE000557 Helicobac
	160	70	1.2	576	2	AF289440	AF289440 Helicobac
	161	70	1.2	10781	1	AE001512	AE001512 Helicobac
	162	69	1.2	291	1	AF198473	AF198473 Helicobac
c	163	69	1.2	2378	2	HPU95957	U95957 Helicobacte
c	164	69	1.2	2963	1	AF190660	AF190660 Helicobac
	165	68	1.1	419	1	AF084492	AF084492 Helicobac
	166	68	1.1	430	1	AF201074	AF201074 Helicobac
	167	68	1.1	589	1	AF191010	AF191010 Helicobac
	168	68	1.1	3524	1	AF249275	AF249275 Helicobac
	169	67	1.1	564	2	AF289444	AF289444 Helicobac
	170	67	1.1	572	2	AF289448	AF289448 Helicobac
	171	67	1.1	576	2	AF289445	AF289445 Helicobac
	172	67	1.1	678	2	AF289459	AF289459 Helicobac
	173	65	1.1	244	2	HPY252965	AJ252965 Helicobac
	174	65	1.1	244	2	HPY252966	AJ252966 Helicobac
	175	65	1.1	513	2	AF289464	AF289464 Helicobac
	176	65	1.1	653	1	AF190662	AF190662 Helicobac
	177	64	1.1	243	2	HPY239720	AJ239720 Helicobac
	178	64	1.1	243	2	HPY239731	AJ239731 Helicobac
	179	62	1.0	286	1	AF198478	AF198478 Helicobac
	180	61	1.0	449	2	HPY269876	AJ269876 Helicobac
	181	61	1.0	12901	1	AE000537	AE000537 Helicobac

c	182	60	1.0	291	2	AF289393	AF289393 Helicobac
c	183	60	1.0	291	2	AF289397	AF289397 Helicobac
	184	59	1.0	464	2	HPY269853	AJ269853 Helicobac
	185	59	1.0	464	2	HPY269874	AJ269874 Helicobac
	186	57	1.0	387	1	AF191007	AF191007 Helicobac
	187	57	1.0	444	1	AF191003	AF191003 Helicobac
	188	57	1.0	513	2	AF289454	AF289454 Helicobac
	189	57	1.0	567	2	AF289452	AF289452 Helicobac
	190	56	0.9	220	1	AF198480	AF198480 Helicobac
	191	56	0.9	286	1	AF202225	AF202225 Helicobac
	192	56	0.9	287	1	AF202219	AF202219 Helicobac
	193	56	0.9	287	1	AF202222	AF202222 Helicobac
	194	56	0.9	452	2	HPY269868	AJ269868 Helicobac
	195	56	0.9	464	2	HPY269856	AJ269856 Helicobac
	196	55	0.9	244	2	HPY252967	AJ252967 Helicobac
	197	55	0.9	427	1	AF201075	AF201075 Helicobac
	198	55	0.9	4735	1	AF049648	AF049648 Helicobac
	199	55	0.9	10482	1	AE000570	AE000570 Helicobac
	200	55	0.9	11148	1	AE001535	AE001535 Helicobac
	201	53	0.9	243	2	HPY239728	AJ239728 Helicobac
	202	52	0.9	413	2	HPY239699	AJ239699 Helicobac
	203	52	0.9	464	2	HPY269870	AJ269870 Helicobac
	204	50	0.8	464	2	HPY269852	AJ269852 Helicobac
	205	49	0.8	204	2	HPCAIDUP	X70038 H.pylori DN
	206	49	0.8	276	2	HPU80066	U80066 Helicobacte
	207	49	0.8	387	1	AF191006	AF191006 Helicobac
	208	49	0.8	413	2	HPY239697	AJ239697 Helicobac
	209	49	0.8	452	2	HPY269857	AJ269857 Helicobac
	210	49	0.8	452	2	HPY269865	AJ269865 Helicobac
	211	49	0.8	464	2	HPY269855	AJ269855 Helicobac
	212	49	0.8	464	2	HPY269858	AJ269858 Helicobac
	213	49	0.8	464	2	HPY269859	AJ269859 Helicobac
	214	49	0.8	464	2	HPY269860	AJ269860 Helicobac
	215	49	0.8	464	2	HPY269861	AJ269861 Helicobac
c	216	48	0.8	291	2	AF289394	AF289394 Helicobac
	217	46	0.8	464	2	HPY269864	AJ269864 Helicobac
	218	46	0.8	511	1	AF200690	AF200690 Helicobac
	219	46	0.8	4724	1	AF049642	AF049642 Helicobac
	220	46	0.8	15374	1	AE000568	AE000568 Helicobac
	221	46	0.8	15549	1	AE001482	AE001482 Helicobac
	222	45	0.8	428	1	AF200689	AF200689 Helicobac
	223	44	0.7	413	2	HPY239700	AJ239700 Helicobac
	224	43	0.7	464	2	HPY269869	AJ269869 Helicobac
	225	42	0.7	452	2	HPY269867	AJ269867 Helicobac
	226	42	0.7	558	1	AF191016	AF191016 Helicobac
	227	41	0.7	640	1	AF190998	AF190998 Helicobac
	228	40	0.7	383	1	AF191005	AF191005 Helicobac
	229	40	0.7	383	1	AF191008	AF191008 Helicobac
	230	40	0.7	464	2	HPY269854	AJ269854 Helicobac
	231	40	0.7	563	1	AF191009	AF191009 Helicobac
	232	40	0.7	629	1	AF190661	AF190661 Helicobac

	233	40	0.7	4751	1	AF049650	AF049650 Helicobac
	234	39	0.7	626	1	AF191011	AF191011 Helicobac
	235	39	0.7	657	1	AF191015	AF191015 Helicobac
	236	38	0.6	413	2	HPY239695	AJ239695 Helicobac
	237	38	0.6	413	2	HPY239696	AJ239696 Helicobac
	238	38	0.6	464	2	HPY269862	AJ269862 Helicobac
	239	38	0.6	464	2	HPY269863	AJ269863 Helicobac
	240	38	0.6	464	2	HPY269871	AJ269871 Helicobac
	241	38	0.6	464	2	HPY269872	AJ269872 Helicobac
	242	38	0.6	464	2	HPY269873	AJ269873 Helicobac
	243	37	0.6	244	2	HPY252972	AJ252972 Helicobac
	244	37	0.6	413	2	HPY239698	AJ239698 Helicobac
	245	36	0.6	263	1	AF198483	AF198483 Helicobac
	246	36	0.6	419	1	AF084494	AF084494 Helicobac
	247	36	0.6	448	1	AF191002	AF191002 Helicobac
	248	35	0.6	243	2	HPY239725	AJ239725 Helicobac
c	249	35	0.6	291	2	AF289390	AF289390 Helicobac
c	250	35	0.6	291	2	AF289391	AF289391 Helicobac
c	251	35	0.6	291	2	AF289395	AF289395 Helicobac
c	252	35	0.6	291	2	AF289396	AF289396 Helicobac
c	253	35	0.6	10700	1	AE000645	AE000645 Helicobac
	254	34	0.6	243	2	HPY239724	AJ239724 Helicobac
	255	34	0.6	464	2	HPY269878	AJ269878 Helicobac
	256	32	0.5	244	2	HPY252975	AJ252975 Helicobac
	257	32	0.5	244	2	HPY252976	AJ252976 Helicobac
	258	31	0.5	378	1	AF191004	AF191004 Helicobac
	259	31	0.5	444	1	AF190999	AF190999 Helicobac
	260	31	0.5	464	2	HPY269877	AJ269877 Helicobac
	261	31	0.5	10818	1	AE000534	AE000534 Helicobac
	262	31	0.5	12216	1	AE000566	AE000566 Helicobac
c	263	31	0.5	12901	1	AE000537	AE000537 Helicobac
c	264	31	0.5	14634	1	AE001442	AE001442 Helicobac
	265	29	0.5	243	2	HPY239727	AJ239727 Helicobac
	266	29	0.5	244	2	HPY252973	AJ252973 Helicobac
	267	29	0.5	244	2	HPY252980	AJ252980 Helicobac
	268	29	0.5	244	2	HPY252984	AJ252984 Helicobac
	269	29	0.5	650	1	AF043487	AF043487 Helicobac
	270	29	0.5	758	1	AF043488	AF043488 Helicobac
	271	28	0.5	18530	1	AE001481	AE001481 Helicobac
	272	27	0.5	243	2	HPY239726	AJ239726 Helicobac
	273	27	0.5	623	1	AF190997	AF190997 Helicobac
	274	26	0.4	244	2	HPY252983	AJ252983 Helicobac
	275	26	0.4	244	2	HPY252985	AJ252985 Helicobac
	276	26	0.4	244	2	HPY252986	AJ252986 Helicobac
	277	26	0.4	14637	1	AE000535	AE000535 Helicobac
	278	25	0.4	243	2	HPY239722	AJ239722 Helicobac
	279	25	0.4	244	2	HPY252977	AJ252977 Helicobac
	280	25	0.4	244	2	HPY252978	AJ252978 Helicobac
	281	25	0.4	244	2	HPY252979	AJ252979 Helicobac
	282	25	0.4	244	2	HPY252981	AJ252981 Helicobac
	283	25	0.4	244	2	HPY252982	AJ252982 Helicobac

284	25	0.4	447	1	AF191000	AF191000 Helicobac
285	25	0.4	447	1	AF191001	AF191001 Helicobac
286	25	0.4	812	1	AF043489	AF043489 Helicobac
287	24	0.4	220	1	AF198484	AF198484 Helicobac
288	24	0.4	220	1	AF198486	AF198486 Helicobac
289	24	0.4	243	2	HPY239733	AJ239733 Helicobac
c 290	24	0.4	291	2	AF289392	AF289392 Helicobac
c 291	24	0.4	291	2	AF289398	AF289398 Helicobac
292	24	0.4	12489	1	AE000600	AE000600 Helicobac
293	24	0.4	21347	2	AF282852	AF282852 Helicobac
c 294	24	0.4	38570	1	AC000108	AC000108 Helicobac
295	24	0.4	321003	84	PFMAL4P3	AL035476 Plasmodiu
296	23	0.4	598	1	AF084489	AF084489 Helicobac
297	23	0.4	598	1	AF084490	AF084490 Helicobac
298	23	0.4	598	1	AF084491	AF084491 Helicobac
c 299	23	0.4	1533	6	AF164385	AF164385 Lithachne
c 300	23	0.4	10218	1	AE001450	AE001450 Helicobac
c 301	23	0.4	10818	1	AE000534	AE000534 Helicobac
c 302	23	0.4	11754	1	AE001503	AE001503 Helicobac
303	23	0.4	154524	38	AC011660	AC011660 Homo sapi
c 304	23	0.4	155775	39	AC011943	AC011943 Homo sapi
305	23	0.4	155788	65	AC079404	AC079404 Homo sapi
c 306	23	0.4	156390	41	AC016918	AC016918 Homo sapi
307	23	0.4	161475	57	AC068633	AC068633 Homo sapi
308	23	0.4	163136	48	HS15D23	AL031599 Human DNA
309	23	0.4	172263	55	AC055832	AC055832 Homo sapi
310	23	0.4	176004	58	AC072041	AC072041 Homo sapi
311	23	0.4	180531	53	AC026344	AC026344 Homo sapi
c 312	23	0.4	195164	39	AC012666	AC012666 Homo sapi
313	23	0.4	199987	53	AC026982	AC026982 Homo sapi
c 314	22	0.4	200	72	HPO249018	AJ249018 Human pol
315	22	0.4	432	2	RPZ82395	Z82395 R.prowazeki
c 316	22	0.4	2100	2	RPTUFRPX	Z54170 R.prowazeki
317	22	0.4	4217	56	TCU77974	U77974 Tribolium c
c 318	22	0.4	8169	40	AC014529	AC014529 Drosophil
319	22	0.4	13611	1	AE001453	AE001453 Helicobac
c 320	22	0.4	33782	56	CEZK909	Z82096 Caenorhabdi
c 321	22	0.4	37599	32	CELK05B2	U29379 Caenorhabdi
322	22	0.4	86617	58	AC069359	AC069359 Homo sapi
323	22	0.4	136769	53	AC026453	AC026453 Homo sapi
c 324	22	0.4	156941	41	AC016664	AC016664 Homo sapi
c 325	22	0.4	157651	66	AL136373	AL136373 Homo sapi
326	22	0.4	162574	55	AC060765	AC060765 Homo sapi
c 327	22	0.4	163371	42	AC019066	AC019066 Homo sapi
328	22	0.4	178075	53	AC027486	AC027486 Homo sapi
c 329	22	0.4	179747	59	AC074237	AC074237 Homo sapi
330	22	0.4	188563	55	AC062026	AC062026 Homo sapi
c 331	22	0.4	190963	51	AC023597	AC023597 Homo sapi
c 332	22	0.4	199489	54	AC027505	AC027505 Homo sapi
333	22	0.4	279110	2	RPXXO3	AJ235272 Rickettsi
334	22	0.4	313965	31	AE003445	AE003445 Drosophil

335	21	0.4	758	1	AF043490	AF043490 Helicobac
c 336	21	0.4	842	79	CNS010D2	AL153366 Anopheles
337	21	0.4	924	47	LJRABIY	Z73936 L.japonicus
338	21	0.4	997	7	ASNFRAG	L35210 Aspergillus
339	21	0.4	1081	7	AX027400	AX027400 Sequence
340	21	0.4	1509	47	SCYGL166W	Z72688 S.cerevisia
c 341	21	0.4	1544	56	DM5SRRNA8	X87880 D.melanogas
342	21	0.4	1604	10	AF041037	AF041037 Homo sapi
343	21	0.4	1620	10	AF035291	AF035291 Homo sapi
344	21	0.4	1968	47	NPENOL	X80474 N.frontalis
345	21	0.4	2372	11	CLCYCBMR	X64588 C.longicaud
346	21	0.4	2520	37	AK026960	AK026960 Homo sapi
347	21	0.4	2554	56	CWU49736	U49736 Corethrella
348	21	0.4	3789	47	SCYGL168W	Z72690 S.cerevisia
c 349	21	0.4	4537	47	YSCPMR1	M25488 Saccharomyc
350	21	0.4	8864	7	CSGTSGGP	X85757 S.cerevisia
c 351	21	0.4	10828	37	AF186772	AF186772 Homo sapi
352	21	0.4	11218	42	AC019998	AC019998 Drosophil
353	21	0.4	12029	31	AE001399	AE001399 Plasmodiu
354	21	0.4	12029	31	AE001427	AE001427 Plasmodiu
c 355	21	0.4	14053	77	HSA224639	AJ224639 Homo sapi
c 356	21	0.4	29564	40	AC014221	AC014221 Drosophil
c 357	21	0.4	30900	8	AC002110	AC002110 Genomic s
358	21	0.4	35364	47	SCCHVII35	Z48518 S.cerevisia
359	21	0.4	36120	32	CELF56D1	U39997 Caenorhabdi
c 360	21	0.4	37909	8	AC002105	AC002105 Genomic s
361	21	0.4	39265	8	AC002107	AC002107 Genomic s
362	21	0.4	41154	78	HSCAMF3X1	X96421 H.sapiens Y
363	21	0.4	54965	40	AC014556	AC014556 Drosophil
c 364	21	0.4	64960	37	AL356379	AL356379 Human DNA
c 365	21	0.4	68727	31	AC004516	AC004516 Drosophil
c 366	21	0.4	71023	29	AC004426	AC004426 Drosophil
c 367	21	0.4	73686	6	AB026649	AB026649 Arabidops
c 368	21	0.4	78598	42	AC019755	AC019755 Drosophil
c 369	21	0.4	78768	29	AC008475	AC008475 Homo sapi
c 370	21	0.4	79296	7	ATAC009176	AC009176 Arabidops
c 371	21	0.4	84440	6	AB010075	AB010075 Arabidops
c 372	21	0.4	87705	78	HSDJ6315	AL049821 Human DNA
373	21	0.4	93553	38	AC010593	AC010593 Homo sapi
c 374	21	0.4	99332	78	HSJ298J18	AL096764 Human DNA
c 375	21	0.4	100000	48	AP000502	AP000502 Homo sapi
c 376	21	0.4	104868	7	ATF15G16	AL132959 Arabidops
377	21	0.4	110000	40	AC013622_1	Continuation (2 of
c 378	21	0.4	111590	68	AL357132	AL357132 Homo sapi
c 379	21	0.4	115407	84	HS1000N6	AL034378 Homo sapi
380	21	0.4	115641	23	ATU95973	U95973 Arabidopsis
381	21	0.4	118086	7	ATF6H11	AL021684 Arabidops
c 382	21	0.4	118271	78	HSJ181J13	AL096801 Human DNA
c 383	21	0.4	118357	48	HS1142C11	AL023877 Human DNA
c 384	21	0.4	120770	38	AC011414	AC011414 Homo sapi
c 385	21	0.4	125062	66	AL139253	AL139253 Homo sapi

386	21	0.4	125350	42	AC020768	AC020768 Homo sapi
387	21	0.4	129541	84	HSAC002099	AC002099 Homo sapi
388	21	0.4	131697	67	AL157943	AL157943 Homo sapi
389	21	0.4	135747	67	AL160167	AL160167 Homo sapi
390	21	0.4	139203	57	AC068311	AC068311 Homo sapi
391	21	0.4	140263	69	AL365194	AL365194 Homo sapi
392	21	0.4	140960	29	AC008491	AC008491 Homo sapi
c 393	21	0.4	142001	7	ATF21P14	AL138642 Arabidops
c 394	21	0.4	144473	58	AC072054	AC072054 Homo sapi
c 395	21	0.4	148065	84	CNS01RHN	AL161871 Homo sapi
396	21	0.4	148142	10	AC011421	AC011421 Homo sapi
c 397	21	0.4	149709	57	AC067727	AC067727 Homo sapi
c 398	21	0.4	149978	29	AC008035	AC008035 Homo sapi
c 399	21	0.4	150183	84	HSDJ614A2	AL121990 Homo sapi
c 400	21	0.4	150331	41	AC016587	AC016587 Homo sapi
c 401	21	0.4	150620	55	AC055124	AC055124 Homo sapi
402	21	0.4	150803	38	AC010228	AC010228 Homo sapi
403	21	0.4	151509	54	AC032009	AC032009 Homo sapi
404	21	0.4	152053	66	AL157714	AL157714 Homo sapi
405	21	0.4	153274	69	AL359708	AL359708 Homo sapi
406	21	0.4	153860	41	AC016301	AC016301 Homo sapi
407	21	0.4	155176	77	HS453P22	Z97987 Human DNA s
c 408	21	0.4	155222	57	AC068416	AC068416 Homo sapi
c 409	21	0.4	155369	37	AL160058	AL160058 Human DNA
410	21	0.4	156267	67	AL161780	AL161780 Homo sapi
c 411	21	0.4	156336	53	AC027320	AC027320 Homo sapi
412	21	0.4	157218	69	AL391097	AL391097 Homo sapi
413	21	0.4	157400	67	AL162717	AL162717 Homo sapi
414	21	0.4	160327	59	AC078813	AC078813 Homo sapi
415	21	0.4	160865	30	AC009818	AC009818 Homo sapi
416	21	0.4	163365	67	AL162212	AL162212 Homo sapi
417	21	0.4	163443	29	AC006280	AC006280 Plasmodiu
418	21	0.4	163631	9	AC009276	AC009276 Homo sapi
419	21	0.4	165392	38	AC011670	AC011670 Homo sapi
c 420	21	0.4	165790	30	AC009809	AC009809 Homo sapi
c 421	21	0.4	166199	51	AC024267	AC024267 Homo sapi
422	21	0.4	167172	10	AC009428	AC009428 Homo sapi
c 423	21	0.4	167437	50	AC023068	AC023068 Homo sapi
424	21	0.4	170201	55	AC055885	AC055885 Homo sapi
c 425	21	0.4	171185	54	AC044781	AC044781 Homo sapi
c 426	21	0.4	171419	40	AC015508	AC015508 Homo sapi
427	21	0.4	171817	54	AC027681	AC027681 Mus muscu
428	21	0.4	172276	52	AC026143	AC026143 Homo sapi
c 429	21	0.4	176213	10	AC009518	AC009518 Human Chr
430	21	0.4	177286	9	AC008085	AC008085 Homo sapi
c 431	21	0.4	177405	69	AL365495	AL365495 Homo sapi
c 432	21	0.4	177738	55	AC055886	AC055886 Homo sapi
c 433	21	0.4	178460	66	AL139040	AL139040 Homo sapi
434	21	0.4	179357	30	AC009625	AC009625 Homo sapi
435	21	0.4	179810	38	AC010546	AC010546 Homo sapi
c 436	21	0.4	180283	10	AF134726	AF134726 Homo sapi

c 437	21	0.4	180483	66	AL137061	AL137061 Homo sapi
c 438	21	0.4	180749	67	AL354983	AL354983 Homo sapi
439	21	0.4	181532	48	CNS01DW2	AL136298 Human chr
440	21	0.4	184078	38	AC010917	AC010917 Drosophil
441	21	0.4	184799	54	AC037448	AC037448 Homo sapi
c 442	21	0.4	185374	84	CNS01BOX	AL109767 Homo sapi
443	21	0.4	186110	48	CNS01BP0	AL109847 Human chr
c 444	21	0.4	186325	59	AC074244	AC074244 Homo sapi
445	21	0.4	186504	9	AC006040	AC006040 Homo sapi
c 446	21	0.4	187239	49	AC021349	AC021349 Homo sapi
447	21	0.4	188017	65	AC079886	AC079886 Rattus no
448	21	0.4	188844	48	CNS01DT0	AL122125 Human chr
449	21	0.4	189231	58	AC073323	AC073323 Homo sapi
450	21	0.4	190072	42	AC019130	AC019130 Homo sapi
c 451	21	0.4	191735	69	AL365185	AL365185 Homo sapi
452	21	0.4	192135	68	AL358492	AL358492 Homo sapi
c 453	21	0.4	192335	66	AL139319	AL139319 Homo sapi
454	21	0.4	193086	66	AP286887	AP286887 Homo sapi
455	21	0.4	194377	49	AC021726	AC021726 Homo sapi
c 456	21	0.4	196149	29	AC004709	AC004709 Plasmodiu
457	21	0.4	196490	29	AC005507	AC005507 Plasmodiu
c 458	21	0.4	196541	42	AC019007	AC019007 Homo sapi
459	21	0.4	196900	49	AC020851	AC020851 Mus muscu
460	21	0.4	197076	69	AL359694	AL359694 Homo sapi
c 461	21	0.4	197915	9	AC005725	AC005725 Homo sapi
462	21	0.4	200463	53	AC026289	AC026289 Homo sapi
c 463	21	0.4	201144	67	AL160290	AL160290 Homo sapi
464	21	0.4	203636	54	AC034142	AC034142 Homo sapi
465	21	0.4	204310	41	AC017094	AC017094 Homo sapi
c 466	21	0.4	210056	38	AC011706	AC011706 Drosophil
c 467	21	0.4	213916	59	AC073473	AC073473 Homo sapi
468	21	0.4	221988	38	AC010999	AC010999 Homo sapi
c 469	21	0.4	234112	56	PFMAL4P2	AL035475 Plasmodiu
c 470	21	0.4	245802	29	AC006279	AC006279 Plasmodiu
471	21	0.4	261083	31	AE003794	AE003794 Drosophil
472	21	0.4	262173	31	AE003808	AE003808 Drosophil
473	21	0.4	286680	32	CEDS4G2A	AC024817 Caenorhab
474	21	0.4	298804	29	AC006911	AC006911 Caenorhab
c 475	21	0.4	298868	31	AE003570	AE003570 Drosophil
476	21	0.4	305018	31	AE003437	AE003437 Drosophil
477	20	0.3	284	81	A39551	A39551 Sequence 39
478	20	0.3	284	81	AR050154	AR050154 Sequence
479	20	0.3	284	81	I23729	I23729 Sequence 50
480	20	0.3	457	47	RSU18556	U18556 Raphanus sa
481	20	0.3	533	6	AF061072	AF061072 Pneumocys
482	20	0.3	609	6	AF126742	AF126742 Zea mays
483	20	0.3	766	6	AF051369	AF051369 Oryza sat
c 484	20	0.3	767	5	BTU27100	U27100 Bacteriopha
485	20	0.3	770	11	S81087	S81087 ob=leptin/O
c 486	20	0.3	918	79	CNS01LQA	AL149955 Anopheles
487	20	0.3	945	47	S65090S2	S65092 lip2=lipase

	488	20	0.3	1170	4	AP077819 AP077819 Cyprinus
	489	20	0.3	1196	11	MMU36238 U36238 Mus musculu
	490	20	0.3	1205	78	HSM802390 AL137626 Homo sapi
	491	20	0.3	1213	7	ATH010819 AJ010819 Arabidopa
	492	20	0.3	1272	6	AF145726 AF145726 Oryza sat
c	493	20	0.3	1309	11	IFNGRMUS4 U69597 Mus musculu
	494	20	0.3	1435	78	HSM800087 AL049302 Homo sapi
	495	20	0.3	1576	20	S54076 S54076 6-phosphofr
	496	20	0.3	1579	4	AF146428 AF146428 Gallus ga
c	497	20	0.3	1666	47	SCYLR018C Z73190 S.cerevisia
c	498	20	0.3	1676	78	HSTGF37 X14891 H.sapiens g
	499	20	0.3	1720	47	SCU85960 U85960 Saccharomyc
	500	20	0.3	1762	77	HSA6198 AJ006198 Homo sapi
	501	20	0.3	1803	78	HSM800752 AL080223 Homo sapi
	502	20	0.3	1804	10	AF007143 AF007143 Homo sapi
	503	20	0.3	1902	37	AK026476 AK026476 Homo sapi
	504	20	0.3	1973	78	HSM802394 AL137629 Homo sapi
	505	20	0.3	2165	8	AB037187 AB037187 Homo sapi
c	506	20	0.3	2169	11	AF264064 AF264064 Mus muscu
	507	20	0.3	2173	81	AR073851 AR073851 Sequence
	508	20	0.3	2173	81	I32035 I32035 Sequence 6
	509	20	0.3	2187	11	MMU88328 U88328 Mus musculu
	510	20	0.3	2187	13	AX022287 AX022287 Sequence
	511	20	0.3	2187	13	AX030754 AX030754 Sequence
	512	20	0.3	2204	4	OTISL1MRD X64885 O.tschawyts
	513	20	0.3	2223	78	HSM800424 AL050124 Homo sapi
	514	20	0.3	2309	81	AR073853 AR073853 Sequence
	515	20	0.3	2309	81	I32037 I32037 Sequence 10
c	516	20	0.3	2476	78	HSL81765 L81765 Homo sapien
c	517	20	0.3	2491	56	DME297734 AJ297734 Drosophil
c	518	20	0.3	2511	32	AF145663 AF145663 Drosophil
c	519	20	0.3	2529	85	HUMTGFB3A J03241 Human trans
c	520	20	0.3	2689	47	SCYLR019W Z73191 S.cerevisia
	521	20	0.3	2692	81	AR073854 AR073854 Sequence
	522	20	0.3	2692	81	I32038 I32038 Sequence 14
	523	20	0.3	2734	81	E04383 E04383 cDNA sequen
c	524	20	0.3	2734	81	E05117 E05117 cDNA sequen
c	525	20	0.3	2830	2	BPXYNB X61495 B. fibrioso
c	526	20	0.3	3003	11	RNDIBGNA Z21846 R.norvegicu
c	527	20	0.3	3016	56	DDU31631 U31631 Dictyosteli
	528	20	0.3	3083	1	AF165314 AF165314 Staphyloc
	529	20	0.3	3228	78	HSM802338 AL137585 Homo sapi
c	530	20	0.3	3510	47	SCYDL241W Z74289 S.cerevisia
c	531	20	0.3	3536	1	AF270099 AF270099 Staphyloc
	532	20	0.3	3784	37	AK025725 AK025725 Homo sapi
	533	20	0.3	3791	10	AF063308 AF063308 Homo sapi
	534	20	0.3	3950	2	BAEAG X99724 B.anthracis
	535	20	0.3	3973	81	AR073855 AR073855 Sequence
	536	20	0.3	3973	81	I32039 I32039 Sequence 21
c	537	20	0.3	4273	47	SCLRG1G X78453 S.cerevisia
c	538	20	0.3	4414	32	BMOGLYTRSY L08106 Silk moth g

c	539	20	0.3	4782	56	GCY17882 Y17882 Geodia cydo
c	540	20	0.3	5396	47	SCYDL240W Z74288 S.cerevisia
	541	20	0.3	5850	56	OGFTBB X79930 O.gibsoni T
	542	20	0.3	6516	11	MMU65742 U65742 Mus musculu
	543	20	0.3	7845	47	SCPPR1GEN X90564 S.cerevisia
c	544	20	0.3	10783	11	RNMFCOABP X96560 R.norvegicu
	545	20	0.3	11534	56	PFAARP1PR Y08926 P.falciparu
c	546	20	0.3	13349	31	AE002781 AE002781 Drosophil
c	547	20	0.3	14553	31	AE001391 AE001391 Plasmodiu
c	548	20	0.3	14796	78	HSU75285 U75285 Homo sapien
	549	20	0.3	16155	1	AE002143 AE002143 Ureaplasm
c	550	20	0.3	16943	1	AE002153 AE002153 Ureaplasm
c	551	20	0.3	19712	2	D78509 D78509 Bacillus su
c	552	20	0.3	23442	37	AL136173 AL136173 Human DNA
c	553	20	0.3	25311	47	SPBC216 AL049558 S.pombe c
c	554	20	0.3	25871	8	AB016625 AB016625 Homo sapi
c	555	20	0.3	29366	47	SC9571X Z49810 S.cerevisia
c	556	20	0.3	31804	40	AC014202 AC014202 Drosophil
c	557	20	0.3	31821	32	CEF58G11 Z81094 Caenorhabdi
c	558	20	0.3	32200	31	AE003176 AE003176 Drosophil
	559	20	0.3	33448	32	CELC34B2 AF043693 Caenorhab
	560	20	0.3	33556	42	AC020474 AC020474 Drosophil
c	561	20	0.3	39486	32	CEF13D12 Z49127 Caenorhabdi
	562	20	0.3	40758	40	AC014583 AC014583 Drosophil
	563	20	0.3	43481	6	AB026648 AB026648 Arabidops
	564	20	0.3	44101	59	AC073960 AC073960 Homo sapi
c	565	20	0.3	47563	41	AC017461 AC017461 Drosophil
	566	20	0.3	48081	40	AC014896 AC014896 Drosophil
c	567	20	0.3	54959	41	AC017215 AC017215 Drosophil
	568	20	0.3	55239	57	AC068220 AC068220 Homo sapi
c	569	20	0.3	57722	84	HSCB33B10 Z82175 Homo sapien
	570	20	0.3	59586	77	HS449017 Z98749 Human DNA s
c	571	20	0.3	61938	51	AC024390 AC024390 Homo sapi
c	572	20	0.3	62238	39	AC012390 AC012390 Drosophil
c	573	20	0.3	65067	40	AC014805 AC014805 Drosophil
	574	20	0.3	65227	44	AC012029 AC012029 Homo sapi
c	575	20	0.3	65643	58	AC073386 AC073386 Homo sapi
	576	20	0.3	67571	85	U51244 U51244 Homo sapien
c	577	20	0.3	67887	39	AC012648 AC012648 Drosophil
c	578	20	0.3	68001	51	AC024011 AC024011 Homo sapi
	579	20	0.3	68200	30	AC009801_3 Continuation (4 of
	580	20	0.3	70591	65	AC079076 AC079076 Homo sapi
	581	20	0.3	71311	66	AC084079 AC084079 Homo sapi
c	582	20	0.3	71311	66	AC084079 AC084079 Homo sapi
c	583	20	0.3	71644	57	AC068157 AC068157 Homo sapi
	584	20	0.3	71751	54	AC034134 AC034134 Homo sapi
	585	20	0.3	72696	59	AC074075 AC074075 Homo sapi
	586	20	0.3	72842	66	AC083957 AC083957 Homo sapi
c	587	20	0.3	72842	66	AC083957 AC083957 Homo sapi
	588	20	0.3	73635	6	AC011915 AC011915 Arabidops
	589	20	0.3	74502	51	AC024300 AC024300 Homo sapi

c	590	20	0.3	74502	51	AC024300 AC024300 Homo sapi
c	591	20	0.3	76458	7	F28K20 AC004793 Arabidops
	592	20	0.3	77694	54	AC037461 AC037461 Homo sapi
	593	20	0.3	79323	31	AC002442 AC002442 Drosophil
	594	20	0.3	79779	78	HSJ710H13 AL121712 Human DNA
c	595	20	0.3	80518	84	PFMAL13PA AL109815 Plasmodiu
	596	20	0.3	81079	41	AC016382 AC016382 Homo sapi
	597	20	0.3	81493	7	AT81KBGEN X98130 A.thaliana
	598	20	0.3	84648	41	AC016461 AC016461 Homo sapi
c	599	20	0.3	85304	77	HS227L5 AL031585 Human DNA
	600	20	0.3	86233	6	AB024034 AB024034 Arabidops
c	601	20	0.3	87021	9	AC006473 AC006473 Homo sapi
	602	20	0.3	89501	10	AF110324 AF110324 Homo sapi
	603	20	0.3	91751	29	AC007973 AC007973 Drosophil
	604	20	0.3	91851	7	ATT16L24 AL138659 Arabidops
	605	20	0.3	92709	40	AC015716 AC015716 Homo sapi
	606	20	0.3	92970	38	AC010692 AC010692 Drosophil
c	607	20	0.3	93121	84	DMBR48A1 AL121844 Drosophil
	608	20	0.3	94236	42	AC020539 AC020539 Drosophil
c	609	20	0.3	94890	10	AC022428 AC022428 Homo sapi
	610	20	0.3	98025	10	AF003530 AF003530 Homo sapi
	611	20	0.3	99083	56	CEY37A1B AL023835 Caenorhab
	612	20	0.3	100000	8	AB020862 AB020862 Homo sapi
c	613	20	0.3	100000	8	AB020862 AB020862 Homo sapi
	614	20	0.3	100000	37	AP000096 AP000096 Homo sapi
	615	20	0.3	100000	48	AP000200 AP000200 Homo sapi
c	616	20	0.3	101115	30	AC008826 AC008826 Homo sapi
	617	20	0.3	101215	9	AC006966 AC006966 Homo sapi
	618	20	0.3	101403	38	AC010663 AC010663 Drosophil
c	619	20	0.3	101826	40	AC015176 AC015176 Drosophil
	620	20	0.3	102991	54	AC034224 AC034224 Homo sapi
c	621	20	0.3	104678	39	AC013423 AC013423 Drosophil
c	622	20	0.3	104871	8	AC004584 AC004584 Homo sapi
	623	20	0.3	104992	29	AC005504 AC005504 Plasmodiu
c	624	20	0.3	107377	6	AC005309 AC005309 Arabidops
	625	20	0.3	107397	69	AP000612 AP000612 Homo sapi
	626	20	0.3	108581	52	AC025458 AC025458 Homo sapi
	627	20	0.3	108969	8	AC004813 AC004813 Homo sapi
c	628	20	0.3	109885	77	HS523G1 AL034375 Human DNA
	629	20	0.3	110000	30	AC010187_0 AC010187 Homo sapi
	630	20	0.3	110000	67	AL353694_1 Continuation (2 of
	631	20	0.3	110000	67	AL353694_2 Continuation (3 of
	632	20	0.3	111178	37	AL136129 AL136129 Human DNA
	633	20	0.3	112027	9	AC007006 AC007006 Homo sapi
c	634	20	0.3	114226	29	AC004710 AC004710 Plasmodiu
c	635	20	0.3	114736	56	PFMAL3P3 Z98547 Plasmodium
	636	20	0.3	115045	8	AC004414 AC004414 Homo sapi
c	637	20	0.3	115809	66	AL136377 AL136377 Homo sapi
	638	20	0.3	115933	65	AC079401 AC079401 Homo sapi
	639	20	0.3	116368	48	HS1112F19 AL034420 Human DNA
c	640	20	0.3	118145	38	AC010664 AC010664 Drosophil

c	641	20	0.3	118595	8	AC002379 AC002379 Human BAC
c	642	20	0.3	119482	41	AC016613 AC016613 Homo sapi
c	643	20	0.3	120379	38	AC011429 AC011429 Homo sapi
	644	20	0.3	120914	41	AC016605 AC016605 Homo sapi
	645	20	0.3	121024	7	CHMPXX X04465 Marchantia
c	646	20	0.3	121224	53	AC026798 AC026798 Homo sapi
	647	20	0.3	121576	30	AC010043 AC010043 Drosophil
c	648	20	0.3	121710	30	AC008917 AC008917 Homo sapi
	649	20	0.3	122099	69	AP000435 AP000435 Homo sapi
c	650	20	0.3	122430	69	AL359770 AL359770 Homo sapi
c	651	20	0.3	122897	70	AP001953 AP001953 Homo sapi
	652	20	0.3	124050	29	AC008199 AC008199 Drosophil
	653	20	0.3	126055	84	HS73J07 AL353999 Homo sapi
c	654	20	0.3	126138	77	HSAY180000 Y180000 Homo sapien
	655	20	0.3	126330	40	AC013745 AC013745 Homo sapi
c	656	20	0.3	126565	30	AC008694 AC008694 Homo sapi
	657	20	0.3	127385	29	AC008467 AC008467 Homo sapi
	658	20	0.3	127506	8	AC005244 AC005244 Homo sapi
	659	20	0.3	128541	39	AC012094 AC012094 Homo sapi
	660	20	0.3	128559	8	AC004677 AC004677 Homo sapi
c	661	20	0.3	128728	37	AL136179 AL136179 Human DNA
	662	20	0.3	129047	7	ATF23K16 AL078620 Arabidops
c	663	20	0.3	129149	77	HS510D11 Z98044 Human DNA s
c	664	20	0.3	129226	77	HS960017 AL022166 Human DNA
c	665	20	0.3	129622	30	AC009353 AC009353 Drosophil
c	666	20	0.3	130351	51	AC024911 AC024911 Homo sapi
c	667	20	0.3	131026	38	AC011510 AC011510 Homo sapi
	668	20	0.3	131314	70	AP000817 AP000817 Homo sapi
	669	20	0.3	131853	41	AC017755 AC017755 Drosophil
	670	20	0.3	131970	29	AC008187 AC008187 Drosophil
	671	20	0.3	132319	9	AC006364 AC006364 Homo sapi
	672	20	0.3	132389	51	AC024292 AC024292 Homo sapi
	673	20	0.3	134515	77	HS354J5 Z95118 Human DNA s
	674	20	0.3	135164	31	AC007082 AC007082 Drosophil
	675	20	0.3	135846	68	AL359257 AL359257 Homo sapi
	676	20	0.3	136418	39	AC012435 AC012435 Homo sapi
	677	20	0.3	139798	59	AC073481 AC073481 Homo sapi
c	678	20	0.3	140210	29	AC002993 AC002993 Homo sapi
c	679	20	0.3	141326	69	AL359709 AL359709 Homo sapi
	680	20	0.3	141787	57	AC068280 AC068280 Homo sapi
c	681	20	0.3	142723	41	AC016517 AC016517 Homo sapi
c	682	20	0.3	143516	49	AC022001 AC022001 Homo sapi
c	683	20	0.3	143800	39	AC011875 AC011875 Homo sapi
c	684	20	0.3	143978	52	AC025028 AC025028 Homo sapi
	685	20	0.3	144104	51	AC023772 AC023772 Homo sapi
	686	20	0.3	144391	67	AL159985 AL159985 Homo sapi
c	687	20	0.3	145491	29	AC008193 AC008193 Drosophil
c	688	20	0.3	145894	54	AC034226 AC034226 Homo sapi
c	689	20	0.3	145963	53	AC027390 AC027390 Homo sapi
	690	20	0.3	146286	69	AL390784 AL390784 Homo sapi
c	691	20	0.3	146863	58	AC073392 AC073392 Oryza sat

c 692	20	0.3	147658	29	AC007718	AC007718 Homo sapi
693	20	0.3	148906	57	AC068051	AC068051 Homo sapi
694	20	0.3	149230	10	AC009947	AC009947 Homo sapi
695	20	0.3	150090	68	AL359096	AL359096 Homo sapi
696	20	0.3	151024	41	AC016855	AC016855 Homo sapi
697	20	0.3	151524	42	AC019321	AC019321 Homo sapi
698	20	0.3	151724	67	AL354943	AL354943 Homo sapi
c 699	20	0.3	152053	50	AC022669	AC022669 Homo sapi
c 700	20	0.3	152181	53	AC027253	AC027253 Homo sapi
701	20	0.3	152192	53	AC027410	AC027410 Mus muscu
702	20	0.3	152210	50	AC022910	AC022910 Homo sapi
c 703	20	0.3	152244	48	AP000352	AP000352 Homo sapi
c 704	20	0.3	152278	69	AL360074	AL360074 Homo sapi
c 705	20	0.3	153289	39	AC013533	AC013533 Homo sapi
706	20	0.3	153404	59	AC073543	AC073543 Homo sapi
c 707	20	0.3	153835	84	CEY6D1	AL023673 Caenorhab
708	20	0.3	154195	40	AC013641	AC013641 Homo sapi
709	20	0.3	154840	54	AC032035	AC032035 Homo sapi
710	20	0.3	154842	68	AL359453	AL359453 Homo sapi
c 711	20	0.3	154848	8	AC002992	AC002992 Homo sapi
c 712	20	0.3	155426	30	AC010162	AC010162 Homo sapi
713	20	0.3	156753	68	AL359175	AL359175 Homo sapi
714	20	0.3	157701	55	AC058814	AC058814 Homo sapi
c 715	20	0.3	158060	29	AC007432	AC007432 Homo sapi
716	20	0.3	158228	54	AC036164	AC036164 Homo sapi
717	20	0.3	158357	68	AL355529	AL355529 Homo sapi
718	20	0.3	158433	53	AC026460	AC026460 Homo sapi
719	20	0.3	158534	84	HSJ492P14	AL121977 Homo sapi
c 720	20	0.3	158831	54	AC027607	AC027607 Homo sapi
c 721	20	0.3	159251	49	AC021659	AC021659 Homo sapi
c 722	20	0.3	159400	48	CNS01DSA	AL121766 Human chr
723	20	0.3	159506	77	HS334I8	AL021918 Homo sapi
724	20	0.3	159635	38	AC011009	AC011009 Homo sapi
c 725	20	0.3	160091	59	AC073991	AC073991 Homo sapi
726	20	0.3	160242	52	AC025928	AC025928 Homo sapi
727	20	0.3	161476	51	AC024697	AC024697 Homo sapi
c 728	20	0.3	161683	40	AC015642	AC015642 Homo sapi
729	20	0.3	161745	38	AC011511	AC011511 Homo sapi
730	20	0.3	161897	55	AC053482	AC053482 Homo sapi
c 731	20	0.3	161935	9	AC008269	AC008269 Homo sapi
c 732	20	0.3	162139	9	AC005682	AC005682 Homo sapi
733	20	0.3	162526	51	AC024182	AC024182 Homo sapi
c 734	20	0.3	162696	70	AP000864	AP000864 Homo sapi
735	20	0.3	163283	65	AC079835	AC079835 Homo sapi
c 736	20	0.3	163778	38	AC010868	AC010868 Homo sapi
c 737	20	0.3	164273	70	AP001542	AP001542 Homo sapi
c 738	20	0.3	164313	51	AC024504	AC024504 Homo sapi
739	20	0.3	164340	39	AC012086	AC012086 Homo sapi
c 740	20	0.3	164451	69	AL442646	AL442646 Homo sapi
741	20	0.3	164646	69	AL359928	AL359928 Homo sapi
742	20	0.3	164830	65	AC079455	AC079455 Homo sapi

c 743	20	0.3	164831	37	AL353765	AL353765 Human DNA
744	20	0.3	164941	57	AC064857	AC064857 Homo sapi
745	20	0.3	164991	29	AC006448	AC006448 Homo sapi
c 746	20	0.3	165242	9	AC009298	AC009298 Homo sapi
c 747	20	0.3	165409	52	AC025335	AC025335 Homo sapi
748	20	0.3	165593	53	AC026856	AC026856 Homo sapi
749	20	0.3	165624	42	AC019037	AC019037 Homo sapi
c 750	20	0.3	165859	40	AC016237	AC016237 Homo sapi
c 751	20	0.3	166053	50	AC023110	AC023110 Homo sapi
c 752	20	0.3	166096	29	AC007910	AC007910 Homo sapi
c 753	20	0.3	166100	58	AC068922	AC068922 Homo sapi
c 754	20	0.3	166126	8	AC004142	AC004142 Homo sapi
755	20	0.3	166215	29	AC007716	AC007716 Homo sapi
c 756	20	0.3	166344	67	AL162577	AL162577 Homo sapi
757	20	0.3	166707	66	AL139116	AL139116 Homo sapi
758	20	0.3	166797	41	AC016718	AC016718 Homo sapi
759	20	0.3	166871	54	AC041046	AC041046 Homo sapi
c 760	20	0.3	166924	69	AL391707	AL391707 Homo sapi
c 761	20	0.3	166989	39	AC013554	AC013554 Homo sapi
762	20	0.3	166998	9	AC006230	AC006230 Homo sapi
763	20	0.3	167133	41	AC016986	AC016986 Homo sapi
c 764	20	0.3	167824	54	AC032025	AC032025 Homo sapi
c 765	20	0.3	168068	48	CNS05TC6	AL355072 Human chr
c 766	20	0.3	168327	58	AC069412	AC069412 Homo sapi
c 767	20	0.3	168365	69	AP000629	AP000629 Homo sapi
768	20	0.3	168464	55	AC053522	AC053522 Homo sapi
769	20	0.3	168651	59	AC074363	AC074363 Homo sapi
c 770	20	0.3	168681	51	AC023670	AC023670 Homo sapi
771	20	0.3	168772	66	AL138961	AL138961 Homo sapi
c 772	20	0.3	168832	70	AP002791	AP002791 Homo sapi
773	20	0.3	168903	5	AP158101	AP158101 Bacteriop
c 774	20	0.3	168963	30	AC009705	AC009705 Homo sapi
c 775	20	0.3	169075	39	AC011940	AC011940 Homo sapi
c 776	20	0.3	169194	52	AC025559	AC025559 Homo sapi
777	20	0.3	169546	29	AC004157	AC004157 Plasmodiu
778	20	0.3	169610	67	AL354854	AL354854 Homo sapi
c 779	20	0.3	169935	84	CNS05TDD	AL356032 Homo sapi
780	20	0.3	169955	30	AC009416	AC009416 Homo sapi
c 781	20	0.3	170138	69	AL391724	AL391724 Homo sapi
782	20	0.3	170637	40	AC015912	AC015912 Homo sapi
783	20	0.3	170796	58	AC069416	AC069416 Homo sapi
784	20	0.3	170917	67	AL161896	AL161896 Homo sapi
785	20	0.3	171034	70	AP001325	AP001325 Homo sapi
c 786	20	0.3	171034	70	AP001325	AP001325 Homo sapi
c 787	20	0.3	171051	51	AC023528	AC023528 Homo sapi
788	20	0.3	171064	55	AC055754	AC055754 Homo sapi
789	20	0.3	171834	10	AC015973	AC015973 Homo sapi
790	20	0.3	171862	30	AC009442	AC009442 Homo sapi
791	20	0.3	172308	53	AC026220	AC026220 Homo sapi
792	20	0.3	172554	51	AC024103	AC024103 Homo sapi
793	20	0.3	172570	9	AC006064	AC006064 Homo sapi

794	20	0.3	172616	37	AL157398	AL157398 Human DNA
c 795	20	0.3	172830	70	AP001320	AP001320 Homo sapi
796	20	0.3	172859	38	AC010815	AC010815 Homo sapi
797	20	0.3	172937	55	AC051642	AC051642 Homo sapi
c 798	20	0.3	172957	41	AC017087	AC017087 Homo sapi
c 799	20	0.3	173230	39	AC012216	AC012216 Homo sapi
800	20	0.3	173242	9	AC006991	AC006991 Homo sapi
c 801	20	0.3	173251	55	AC058798	AC058798 Homo sapi
802	20	0.3	173314	10	AC016637	AC016637 Homo sapi
803	20	0.3	173521	57	AC067850	AC067850 Homo sapi
804	20	0.3	173654	70	AP001269	AP001269 Homo sapi
805	20	0.3	173846	59	AC073643	AC073643 Homo sapi
c 806	20	0.3	173907	10	AC022313	AC022313 Homo sapi
807	20	0.3	173949	51	AC023566	AC023566 Homo sapi
808	20	0.3	174345	52	AC025236	AC025236 Homo sapi
c 809	20	0.3	174368	29	AC006247	AC006247 Drosophil
810	20	0.3	174416	42	AC018898	AC018898 Homo sapi
811	20	0.3	174515	52	AC025944	AC025944 Homo sapi
c 812	20	0.3	174533	59	AC073718	AC073718 Mus muscu
813	20	0.3	174625	58	AC073159	AC073159 Homo sapi
814	20	0.3	174677	29	AC008049	AC008049 Homo sapi
c 815	20	0.3	175300	39	AC012459	AC012459 Homo sapi
c 816	20	0.3	175653	30	AC009409	AC009409 Homo sapi
c 817	20	0.3	175668	53	AC027268	AC027268 Homo sapi
818	20	0.3	176012	39	AC011795	AC011795 Homo sapi
c 819	20	0.3	176099	48	AP000240	AP000240 Homo sapi
c 820	20	0.3	176146	50	AC023124	AC023124 Homo sapi
821	20	0.3	176185	84	CNS01RI0	AL162791 Homo sapi
822	20	0.3	176261	68	AL356775	AL356775 Homo sapi
c 823	20	0.3	176279	57	AC066688	AC066688 Mus muscu
824	20	0.3	176340	53	AC026361	AC026361 Homo sapi
c 825	20	0.3	176666	52	AC025611	AC025611 Homo sapi
826	20	0.3	176742	9	AC007386	AC007386 Homo sapi
c 827	20	0.3	176745	39	AC011958	AC011958 Homo sapi
828	20	0.3	177215	53	AC026886	AC026886 Homo sapi
829	20	0.3	177350	84	HSDJ90G1	AL132670 Homo sapi
830	20	0.3	177362	67	AL158154	AL158154 Homo sapi
831	20	0.3	177663	59	AC074129	AC074129 Homo sapi
c 832	20	0.3	177663	59	AC074129	AC074129 Homo sapi
c 833	20	0.3	177675	48	CNS01DT5	AL132640 Human chr
c 834	20	0.3	177808	84	CNS05TD9	AL356019 Homo sapi
835	20	0.3	177899	59	AC074175	AC074175 Mus muscu
c 836	20	0.3	177924	54	AC036141	AC036141 Homo sapi
837	20	0.3	178019	70	AP002499	AP002499 Homo sapi
c 838	20	0.3	178042	51	AC023986	AC023986 Homo sapi
c 839	20	0.3	178500	69	AL359926	AL359926 Homo sapi
c 840	20	0.3	178607	9	AC007068	AC007068 Homo sapi
841	20	0.3	178629	66	AL138956	AL138956 Homo sapi
c 842	20	0.3	178710	10	AC011477	AC011477 Homo sapi
843	20	0.3	178794	54	AC040992	AC040992 Homo sapi
c 844	20	0.3	178799	68	AL357752	AL357752 Homo sapi

845	20	0.3	178861	57	AC068231	AC068231 Homo sapi
c 846	20	0.3	179454	51	AC024177	AC024177 Homo sapi
c 847	20	0.3	179638	42	AC020675	AC020675 Homo sapi
848	20	0.3	179771	42	AC019090	AC019090 Homo sapi
849	20	0.3	179814	54	AC036232	AC036232 Homo sapi
850	20	0.3	180014	44	AC012015	AC012015 Homo sapi
851	20	0.3	180256	52	AC025212	AC025212 Homo sapi
852	20	0.3	180279	67	AL353775	AL353775 Homo sapi
c 853	20	0.3	180379	66	AL137249	AL137249 Homo sapi
c 854	20	0.3	180493	38	AC011189	AC011189 Homo sapi
c 855	20	0.3	181611	70	AP000848	AP000848 Homo sapi
c 856	20	0.3	182064	66	AC083868	AC083868 Homo sapi
c 857	20	0.3	182082	41	AC016583	AC016583 Homo sapi
858	20	0.3	183077	50	AC022193	AC022193 Homo sapi
c 859	20	0.3	183128	41	AC016768	AC016768 Homo sapi
860	20	0.3	183170	50	AC022146	AC022146 Homo sapi
861	20	0.3	183713	49	AC021074	AC021074 Homo sapi
862	20	0.3	183721	67	AL158826	AL158826 Homo sapi
c 863	20	0.3	183849	65	AC083751	AC083751 Oryza sat
864	20	0.3	184506	30	AC009448	AC009448 Homo sapi
c 865	20	0.3	184535	39	AC012636	AC012636 Homo sapi
866	20	0.3	184656	58	AC073150	AC073150 Homo sapi
867	20	0.3	184852	52	AC026032	AC026032 Homo sapi
868	20	0.3	185149	51	AC023800	AC023800 Homo sapi
869	20	0.3	185203	67	AL161649	AL161649 Homo sapi
870	20	0.3	185475	40	AC015846	AC015846 Homo sapi
c 871	20	0.3	185604	40	AC015801	AC015801 Homo sapi
c 872	20	0.3	185922	39	AC012574	AC012574 Homo sapi
873	20	0.3	186509	67	AL162393	AL162393 Homo sapi
c 874	20	0.3	186587	68	AL355380	AL355380 Homo sapi
875	20	0.3	186930	49	AC021494	AC021494 Homo sapi
876	20	0.3	187075	49	AC021617	AC021617 Homo sapi
877	20	0.3	187079	68	AL359207	AL359207 Homo sapi
c 878	20	0.3	187256	29	AC008272	AC008272 Homo sapi
879	20	0.3	187279	53	AC026698	AC026698 Homo sapi
c 880	20	0.3	187903	66	AL135903	AL135903 Homo sapi
c 881	20	0.3	188462	58	AC073051	AC073051 Homo sapi
c 882	20	0.3	188781	50	AC022546	AC022546 Homo sapi
883	20	0.3	188948	9	AC005832	AC005832 Homo sapi
c 884	20	0.3	189219	30	AC009680	AC009680 Homo sapi
885	20	0.3	189764	70	AP000897	AP000897 Homo sapi
c 886	20	0.3	190075	70	AP000779	AP000779 Homo sapi
c 887	20	0.3	190259	66	AL139111	AL139111 Homo sapi
888	20	0.3	190336	57	AC068615	AC068615 Homo sapi
889	20	0.3	190390	31	AC005720	AC005720 Drosophil
c 890	20	0.3	190719	70	AP001561	AP001561 Homo sapi
891	20	0.3	190976	67	AL158134	AL158134 Homo sapi
892	20	0.3	191145	9	AC006112	AC006112 Homo sapi
c 893	20	0.3	191356	10	AC009455	AC009455 Homo sapi
894	20	0.3	191397	51	AC023861	AC023861 Homo sapi
895	20	0.3	191425	66	AL137850	AL137850 Homo sapi

896	20	0.3	191796	39	AC012555	AC012555 Homo sapi
c 897	20	0.3	191998	59	AC073869	AC073869 Homo sapi
898	20	0.3	192126	48	DJ270M14	AP107885 Homo sapi
899	20	0.3	193218	57	AC068605	AC068605 Mus muscu
c 900	20	0.3	193487	70	AP001793	AP001793 Homo sapi
c 901	20	0.3	193999	55	AC048361	AC048361 Mus muscu
902	20	0.3	195294	55	AC051623	AC051623 Mus muscu
903	20	0.3	195694	51	AC024115	AC024115 Mus muscu
904	20	0.3	196247	84	HSJ651N20	AL117346 Homo sapi
905	20	0.3	196403	54	AC027767	AC027767 Mus muscu
906	20	0.3	196416	9	AC006483	AC006483 Homo sapi
c 907	20	0.3	196490	29	AC005507	AC005507 Plasmodiu
c 908	20	0.3	197176	50	AC022763	AC022763 Homo sapi
909	20	0.3	198151	7	ATCHRIV91	AL161595 Arabidops
c 910	20	0.3	199665	10	AP064857	AP064857 Homo sapi
911	20	0.3	199854	58	AC069226	AC069226 Homo sapi
912	20	0.3	199882	37	AL354720	AL354720 Human DNA
c 913	20	0.3	200567	70	AP001587	AP001587 Homo sapi
914	20	0.3	200902	68	AL356599	AL356599 Homo sapi
915	20	0.3	201020	51	AC024067	AC024067 Homo sapi
c 916	20	0.3	201371	59	AC073786	AC073786 Mus muscu
c 917	20	0.3	202013	51	AC023765	AC023765 Homo sapi
918	20	0.3	202405	40	AC015884	AC015884 Homo sapi
919	20	0.3	202712	40	AC015676	AC015676 Homo sapi
920	20	0.3	204190	50	AC023206	AC023206 Homo sapi
921	20	0.3	204504	41	AC016820	AC016820 Homo sapi
922	20	0.3	204720	29	AC008052	AC008052 Homo sapi
c 923	20	0.3	207152	55	AC062004	AC062004 Homo sapi
c 924	20	0.3	207389	41	AC017115	AC017115 Homo sapi
c 925	20	0.3	208424	40	AC013722	AC013722 Homo sapi
926	20	0.3	208430	2	BSUB0005	Z99108 Bacillus su
927	20	0.3	208642	77	HSA508N22	AL117339 Human DNA
c 928	20	0.3	208684	58	AC072048	AC072048 Mus muscu
c 929	20	0.3	209060	42	AC019005	AC019005 Homo sapi
c 930	20	0.3	209861	49	AC021196	AC021196 Homo sapi
931	20	0.3	210008	41	AC016974	AC016974 Homo sapi
c 932	20	0.3	210051	69	AL442070	AL442070 Homo sapi
c 933	20	0.3	213475	10	AC019100	AC019100 Homo sapi
934	20	0.3	214186	65	AC078933	AC078933 Mus muscu
935	20	0.3	215447	68	AL357147	AL357147 Homo sapi
936	20	0.3	217608	58	AC073087	AC073087 Homo sapi
937	20	0.3	218270	65	AC079499	AC079499 Mus muscu
c 938	20	0.3	219194	49	AC022061	AC022061 Mus muscu
c 939	20	0.3	223890	55	AC062008	AC062008 Homo sapi
c 940	20	0.3	226345	8	AC005406	AC005406 Homo sapi
941	20	0.3	226383	68	AL355593	AL355593 Homo sapi
942	20	0.3	227722	59	AC073712	AC073712 Mus muscu
943	20	0.3	227996	68	AL355859	AL355859 Homo sapi
c 944	20	0.3	229426	38	AC010532	AC010532 Homo sapi
945	20	0.3	232147	41	AC018467	AC018467 Homo sapi
c 946	20	0.3	232156	31	AE003819	AE003819 Drosophil

c 947	20	0.3	233313	31	AE003741	AE003741 Drosophil
c 948	20	0.3	234145	40	AC015687	AC015687 Homo sapi
949	20	0.3	234400	58	AC069254	AC069254 Homo sapi
950	20	0.3	240000	29	AC006393	AC006393 Homo sapi
c 951	20	0.3	241359	68	AL355585	AL355585 Homo sapi
c 952	20	0.3	242396	31	AE003599	AE003599 Drosophil
953	20	0.3	245802	29	AC006279	AC006279 Plasmodiu
c 954	20	0.3	249287	29	AC006745	AC006745 Caenorhab
c 955	20	0.3	261486	31	AE003827	AE003827 Drosophil
956	20	0.3	263704	31	AE003667	AE003667 Drosophil
957	20	0.3	265605	31	AE003662	AE003662 Drosophil
958	20	0.3	268960	65	AC078972	AC078972 Homo sapi
c 959	20	0.3	269518	31	AE003751	AE003751 Drosophil
960	20	0.3	274307	31	AE003561	AE003561 Drosophil
c 961	20	0.3	278229	59	AC074322	AC074322 Homo sapi
962	20	0.3	281869	31	AE003544	AE003544 Drosophil
c 963	20	0.3	289393	68	AL355877	AL355877 Homo sapi
c 964	20	0.3	293431	84	PFMAL13P4	AL049181 Plasmodiu
965	20	0.3	298640	31	AE003467	AE003467 Drosophil
c 966	20	0.3	300710	31	AE003444	AE003444 Drosophil
967	20	0.3	302915	31	AE003489	AE003489 Drosophil
c 968	20	0.3	306464	57	AC068708	AC068708 Homo sapi
c 969	20	0.3	310565	84	CEY62F5	AL023633 Caenorhab
970	20	0.3	316170	84	CEY37A1	Z93243 Caenorhabdi
971	20	0.3	316758	31	AE003786	AE003786 Drosophil
972	20	0.3	334796	53	AC026673	AC026673 Homo sapi
c 973	20	0.3	337101	78	HSXKSRPXR	AL121578 Homo sapi
c 974	20	0.3	340000	48	AP001705	AP001705 Homo sapi
c 975	20	0.3	340000	77	HS21C081	AL163281 Homo sapi
c 976	19	0.3	35	81	AR051296	AR051296 Sequence
c 977	19	0.3	35	81	I16944	I16944 Sequence 13
c 978	19	0.3	35	81	I45738	I45738 Sequence 13
c 979	19	0.3	124	56	DDIACTJ	M29118 D.discoideu
980	19	0.3	126	79	G37840	G37840 Send Plasmo
981	19	0.3	220	1	AF198485	AF198485 Helicobac
c 982	19	0.3	224	4	AF250526	AF250526 Brycon mi
c 983	19	0.3	224	4	AF250527	AF250527 Brycon mi
c 984	19	0.3	273	80	HSPF16E2	Z94668 H.sapiens f
985	19	0.3	322	32	AF231991	AF231991 Sterkiell
986	19	0.3	419	1	AF084493	AF084493 Helicobac
c 987	19	0.3	450	79	G22206	G22206 human STS W
c 988	19	0.3	484	32	AF097151	AF097151 Chelymorp
c 989	19	0.3	504	80	G61216	G61216 SHGC-85524
990	19	0.3	570	47	PUPSAE	X60434 P.purpurea
991	19	0.3	584	7	AF233529	AF233529 Penicilli
992	19	0.3	645	78	HSMB02774	AL390165 Homo sapi
993	19	0.3	650	54	AC029231	AC029231 Giardia i
994	19	0.3	660	47	YSCSNR128	M21124 S.cerevisia
995	19	0.3	669	31	AB026847	AB026847 Ciona int
996	19	0.3	671	56	CIN249323	AJ249323 Ciona int
c 997	19	0.3	736	79	G52975	G52975 SHGC-86334

998	19	0.3	760	48	AX011739	AX011739 Sequence
999	19	0.3	778	47	PAU66076	U66076 Prunus aviu
c1000	19	0.3	780	79	CNS01KMO	AL148529 Anopheles

ALIGNMENTS

```
RESULT   1
AX030324
LOCUS       AX030324      5925 bp    DNA              BCT       16-SEP-2000
DEFINITION  Sequence 3 from Patent EP0967279.
ACCESSION   AX030324
VERSION     AX030324.1  GI:10190490
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 5925)
  AUTHORS   Covacci,A., Rappuoli,R., Bugnoli,M., Telford,J. and Macchia,G.
  TITLE     Helicobacter pylori cytotoxin useful for vaccines and diagnostics
  JOURNAL   Patent: EP 0967279-A 29-DEC-1999;
            CHIRON SPA (IT)
FEATURES             Location/Qualifiers
     source          1..5925
                     /organism="Helicobacter pylori"
                     /db_xref="taxon:210"
BASE COUNT     2076 a    1057 c    1138 g    1654 t
ORIGIN
```

Query Match 100.0%; Score 5925; DB 2; Length 5925;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5925; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ctccatttttaagcaactccatagaccactaaagaaactttttttgaggctatctttgaaa 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCCATTTTAAGCAACTCCATAGACCACTAAAGAAACTTTTTTTGAGGCTATCTTTGAAA 60

Qy      61 atctgtcctattgatttgttttccatttttgtttcccatgtggatcttgtggatcacaaac 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCTGTCCTATTGATTTGTTTTCCATTTTGTTTCCCATGTGGATCTTGTGGATCACAAAC 120

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTTAATTATACATGCTATAGTAAGCATGACACACAAACCAAACTATTTTTAGAACGCTT 180

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgattttct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CATGTGCTCACCTTGACTAACCATTTCTCCAACCATACTTTAGCGTTGCATTTGATTTCT 240
```



```
Qy     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttttagcaaatttt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCAAAAAGATTCATTTCTTATTTCTTGTTCTTATTAAAGTTCTTTCATTTTAGCAAATTT 300

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTGTTAATTGTGGGTAAAAATGTGAATCGTCCTAGCCTTTAGACGCCTGCAACGATCGGG 360

Qy     361 ctttttttcaatattaataatgattaatgaaaaaaaaaaaaaaatgcttgatattgttgtat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTTTTTTCAATATTAATAATGATTAATGAAAAAAAAAAAAAAATGCTTGATATTGTTGTAT 420

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AATGAGAATGTTCAAAGACATGAATTGACTACTCAAGCGTGTAGCGATTTTTAGCAGTCT 480

Qy     481 ttgacactaacaagatacgatagggtatgaaactaggtatagtaaggagaaacaatgact 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTGACACTAACAAGATACCGATAGGTATGAAACTAGGTATAGTAAGGAGAAACAATGACT 540

Qy     541 aacgaaaccattgaccaacaaccacaaaccgaagcggctttaacccgcagcaatttatc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AACGAAACCATTGACCAACAACCACAAACCGAAGCGGCTTTAACCCGCAGCAATTTATC 600

Qy     601 aataatcttcaagtagctttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGAT 660

Qy     661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAAGGAATCTCG 720

Qy     721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTAT 780

Qy     781 ttttcagacttttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTA 840

Qy     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGT 900

Qy     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatg 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATG 960

Qy     961 gaaaatatcatacaaccccctatccttgatgataaagagaaagcggagtttttgaaatct 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    961 GAAAATATCATACAACCCCCTATCCTTGATGATAAAGAGAAAGCGGAGTTTTTGAAATCT 1020

Qy   1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCCAAACAATCTTTTGCAGGAATCATTATAGGGAATCAAATCCGAACGGATCAAAAGTTC 1080

Qy   1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ATGGGCGTGTTTGATGAGTCCTTGAAAGAAAGGCAAGAAGCAGAAAAAAATGGAGAGCCT 1140

Qy   1141 actggtgggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACTGGTGGGGATTGGTTGGATATTTTTCTCTCATTTATATTTGACAAAAAACAATCTTCT 1200

Qy   1201 gatgtccaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GATGTCAAAGAAGCAATCAATCAAGAACCAGTTCCCCATGTCCAACCAGATATAGCCACT 1260

Qy   1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ACCACCACCGACATACAAGGCTTACCGCCTGAAGCTAGAGATTTACTTGATGAAAGGGGT 1320

Qy   1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AATTTTTCTAAATTCACTCTTGGCGATATGGAAATGTTAGATGTTGAGGGAGTCGCTGAC 1380

Qy   1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ATTGATCCCAATTACAAGTTCAATCAATTATTGATTCACAATAACGCTCTGTCTTCTGTG 1440

Qy   1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TTAATGGGGAGTCATAATGGCATAGAACCTGAAAAAGTTTCATTGTTGTATGGGGGCAAT 1500

Qy   1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GGTGGTCCTGGAGCTAGGCATGATTGGAACGCCACCGTTGGTTATAAAGACCAACAAGGC 1560

Qy   1561 aacaatgtggctacaatattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 1620

Qy   1621 ggtggtgagaaagggattaacaaccctagttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 1680

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 1740

Qy   1741 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCGAGAAAGAGAAGGAAAAATTCCGA 1800

Qy   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ACTGAGATTAAAGATTTCCAAAAAGACTCTAAGGCTTATTTAGACGCCCTAGGGAATGAT 1860

Qy   1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CGTATTGCTTTTGTTTCTAAAAAAGACACAAAACATTCAGCTTTAATTACTGAGTTTGGT 1920

Qy   1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AATGGGGATTTGAGCTACACTCTCAAAGATTATGGGAAAAAGCAGATAAAGCTTTAGAT 1980

Qy   1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AGGGAGAAAAATGTTACTCTTCAAGGTAGCCTAAAACATGATGGCGTGATGTTTGTTGAT 2040

Qy   2041 tattctaatttcaaatacaccaacgcctccaagaatcccaatgaaggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TATTCTAATTTCAAATACACCAACGCCTCCAAGAATCCCAATGAAGGGTGTAGGCGTTACG 2100

Qy   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AATGGCGTTTCCCATTTAGAAGTAGGCTTTAACAAGGTAGCTATCTTTAATTTGCCTGAT 2160

Qy   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TTAAATAATCTCGCTATCACTAGTTTCGTAAGGCGGAATTTAGAGGATAAACTAACCACT 2220

Qy   2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttggagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AAAGGATTGTCCCCACAAGAAGCTAATAAGCTTATCAAAGATTTTTGGAGCAGCAACAAA 2280

Qy   2281 gaattggttggaaaaacttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GAATTGGTTGGAAAAACTTTAAACTTCAATAAAGCTGTAGCTGACGCTAAAAACACAGGC 2340

Qy   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 AATTATGATGAAGTGAAAAAAGCTCAGAAAGATCTTGAAAAATCTCTAAGGAAACGAGAG 2400

Qy   2401 catttagagaaagaagtagagaaaaaattggagagcgcaaaagcggcaacaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CATTTAGAGAAAGAAGTAGAGAAAAAATTGGAGAGCAAAAGCGGCAACAAAAATAAAATG 2460

Qy   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GAAGCAAAAGCTCAAGCTAACAGCCAAAAAGATGAGATTTTTGCGTTGATCAATAAAGAG 2520
```



```
Qy   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GCTAATAGAGACGCAAGAGCAATCGCTTACGCTCAGAATCTTAAAGGCATCAAAAGGGAA 2580

Qy   2581 ttgtctgataaacttgaaaatgtcaaccaagaatttgaaagactttgataaatctttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TTGTCTGATAAACTTGAAAATGTCAACAAGAATTTGAAAGACTTTGATAAATCTTTTGAT 2640

Qy   2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagccctt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 GAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGCAGAAGAAACACTAAAAGCCCTT 2700

Qy   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaacctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 AAAGGTTCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAAAGTTGAAAACCTT 2760

Qy   2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 AATGCAGCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGTAACGCAA 2820

Qy   2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GCAAAAAGCGACCTTGAAAATTCCGTTAAAGATGTGATCATCAATCAAAAGGTAACGGAT 2880

Qy   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 AAAGTTGATAATCTCAATCAAGCGGTATCAGTGGCTAAAGCAACGGGTGATTTCAGTAGG 2940

Qy   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaacaagct 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 GTAGAGCAAGCGTTAGCCGATCTCAAAAATTTCTCAAAGGAGCAATTGGCCCAACAAGCT 3000

Qy   3001 caaaaaaatgaaagtctcaatgctagaaaaaaaatctgaaatatatcaatccgttaagaat 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 CAAAAAAATGAAAGTCTCAATGCTAGAAAAAAATCTGAAATATATCAATCCGTTAAGAAT 3060

Qy   3061 ggtgtgaatggaacccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 GGTGTGAATGGAACCCTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACTCTTTCT 3120

Qy   3121 aaaaactttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAAAACTTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAAT 3180

Qy   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AACAATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGG 3240

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 CAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAA 3300
```



```
Qy   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 ATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC 3360

Qy   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAA 3420

Qy   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTT 3480

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATG 3540

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 AATCTATGGGTTGAAAGTGCAAAAAAAGTACCTGCTAGTTTGTCAGCGAAACTAGACAAT 3600

Qy   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 TACGCTACTAACAGCCACATACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAA 3660

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 AAAGCGACCGGCATGCTAACGCAAAAAAACCCTGAGTGGCTCAAGCTCGTGAATGATAAG 3720

Qy   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 ATAGTTGCGCATAATGTAGGAAGCGTTCCTTTGTCAGAGTATGATAAAATTGGCTTCAAC 3780

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 CAGAAGAATATGAAAGATTATTCTGATTCGTTCAAGTTTTCCACCAAGTTGAACAATGCT 3840

Qy   3841 gtaaaagacactaatttctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 GTAAAAGACACTAATTCTGGCTTTACGCAATTTTTAACCAATGCATTTTCTACAGCATCT 3900

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TATTACTGCTTGGCGAGAGAAAATGCGGAGCATGGAATCAAGAACGTTAATACAAAAGGT 3960

Qy   3961 ggtttccaaaaatcttaaaggattaaggaataccaaaaacgcaaaaaccacccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GGTTTCCAAAAATCTTAAAGGATTAAGGAATACCAAAAACGCAAAAACCACCCCTTGCTA 4020

Qy   4021 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   4021 AAAGCGAGGGGTTTTTTAATACTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGA 4080

Qy   4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 TGAATGCTACCAATTCATGGTATCATATCCCCATACATTCGTATCTAGCGTAGGAAGTGT 4140

Qy   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 4200

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 AACTCTGTAAAATCCCTATTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 4260

Qy   4261 acatcagcctagaaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 4320

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAA 4380

Qy   4381 aagacttaacccttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 AAGACTTAACCCTTTGCTTAAAATTAAGTTTGATTGTGCTAGTGGGTTCGTGCTATAGTG 4440

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 CGAAAATTAATTAAGGGTTATAAAGAGAGCATAAACTAGAAAAAACAAGTAGCTATAACA 4500

Qy   4501 aagatcaagttcaaaaaaatcatagagcttttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 AAGATCAAGTTCAAAAAATCATAGAGCTTTTAGAGCAAATTGATCGCGCTCTTAACCAAA 4560

Qy   4561 gaaaaatcagaaaaaccatagaattatcacaccttataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 GAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAAGACGCTTGC 4620

Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 GATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACTGTGG 4680

Qy   4681 atgccttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 ATGCCTTCAAGGTGAAGAGGCAGATATTATTATTTATTCCACCGTGAAAACTTGTGGTA 4740

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 ATCTTTCTTTCTTGCTAGATTCTAAACGCTTGAATGTGGCTATTTCTAGGGCAAAAGAAA 4800

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 ATCTCATTTTTGTGGGTAAAAAGTCTTTCTTTGAGAATTTATGAAGCGATGAGAAGAATA 4860

Qy   4861 tctttagcgctatttttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 TCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 4920

Qy   4921 cattcttcgctttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACACAAT 4980

Qy   4981 acccttatagtgtgagctatagccccttttttgggaattgagttattttgacttttaaattt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 ACCCTTATAGTGTGAGCTATAGCCCCTTTTTGGGAATTGAGTTATTTTGACTTTAAATTT 5040

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCC 5100

Qy   5101 gctcgcatgaaattccacctttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 GCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATTTTTTTTAAGGGCGTATTTTTG 5160

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggggtgttgaaagggc 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 CTGCAAATATCCTACAATAGCATCGCCCGAATGGATGAGTAGGGGGGGGTGTTGAAAGGGC 5220

Qy   5221 aaaatgctccataaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 AAAATGCTCCATAAAATAGCCCTCAATTTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 5280

Qy   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 TAAAATAATCACTTCGGGAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTTTCT 5340

Qy   5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 AACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAAGAGAAGTGGCTAAA 5400

Qy   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 TGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGGATC 5460

Qy   5461 gtcgcttttgtccctagcactaaaatagggggcgtttttatcttttacttgtcgcttgatc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 GTCGCTTTTGTCCCTAGCACTAAAATAGGGGCGTTTTTATCTTTTACTTGTCGCTTGATC 5520

Qy   5521 gctaaaatgcttggctcaatcacgcccacaatagggatttttggaatgcttttgcatctct 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 GCTAAAATGCTTGGCTCAATCACGCCCACAATAGGGATTTTGGAATGCTTTTGCATCTCT 5580
```



```
Qy   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 TCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGGT 5640

Qy   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 TTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTGCCA 5700

Qy   5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 TAAGGCACTCTAGCCGTATCGCCATAATAGATGATTTCATCAAATAATTGCGCTTTTAAA 5760

Qy   5761 aggctttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctattttcatgaca 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 AGGCTTTTTAAAACGCTAAACCCTCCCACACCGCTATCAAAAACGCCTATTTTCATGACA 5820

Qy   5821 ctttttaatttaatgggattaattagggatttttatttttcattcattaagtttaaaaat 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 CTTTTTTAATTTAATGGGATTAATTAGGGATTTTATTTTTCATTCATTAAGTTTAAAAAT 5880

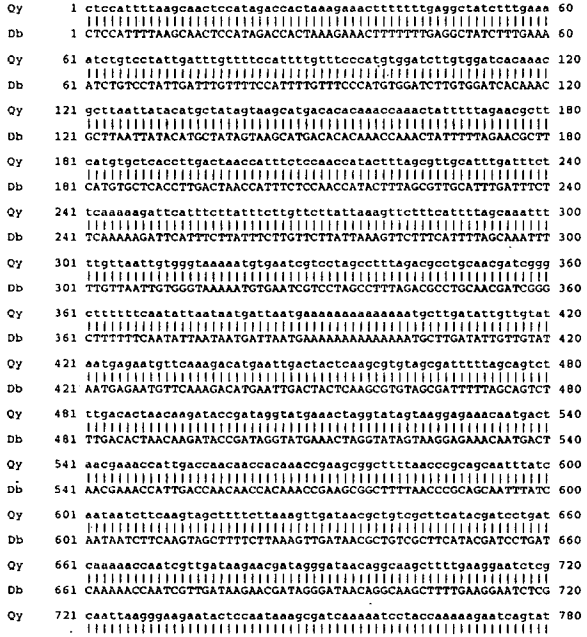
Qy   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
          |||||||||||||||||||||||||||||||||||||||||||||
Db   5881 TCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATAGACAAAGCTT 5925

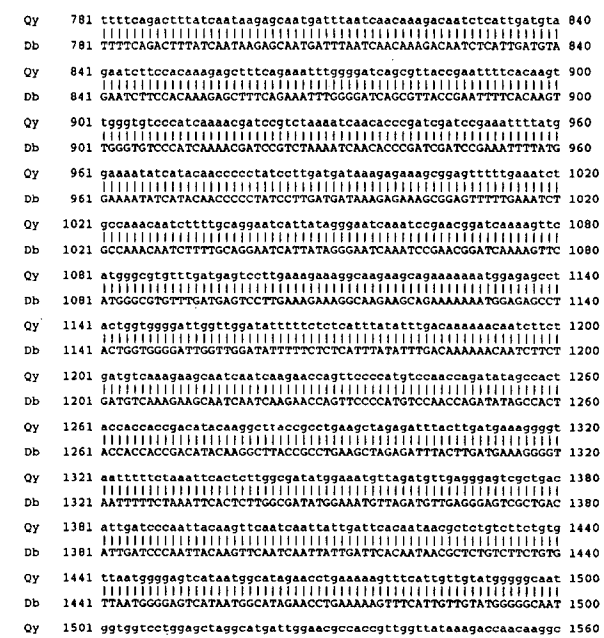

RESULT   2
HPCAI
LOCUS       HPCAI        5925 bp    DNA             BCT       22-JUL-1993
DEFINITION  H.pylori cai gene for cytotoxicity associated immunodominant
            antigen.
ACCESSION   X70039
VERSION     X70039.1  GI:394912
KEYWORDS    cytotoxicity associated immunodominant antigen.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 5925)
  AUTHORS   Covacci,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1993) A. Covacci, Iris, Immunobiological Research
            Inst Siena, Via Fiorentina, 1, 53100 Siena, ITALY
REFERENCE   2  (bases 1 to 5925)
  AUTHORS   Covacci,A., Censini,S., Bugnoli,M., Petracca,R., Burroni,D.,
            Macchia,G., Massone,A., Papini,E., Xiang,Z., Figura,N. and
            Rappuoli,R.
  TITLE     Molecular characterization of the 128-kDa immunodominant antigen of
            Helicobacter pylori associated with cytotoxicity and duodenal ulcer
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90 (12), 5791-5795 (1993)
  MEDLINE   93296225
FEATURES             Location/Qualifiers
```

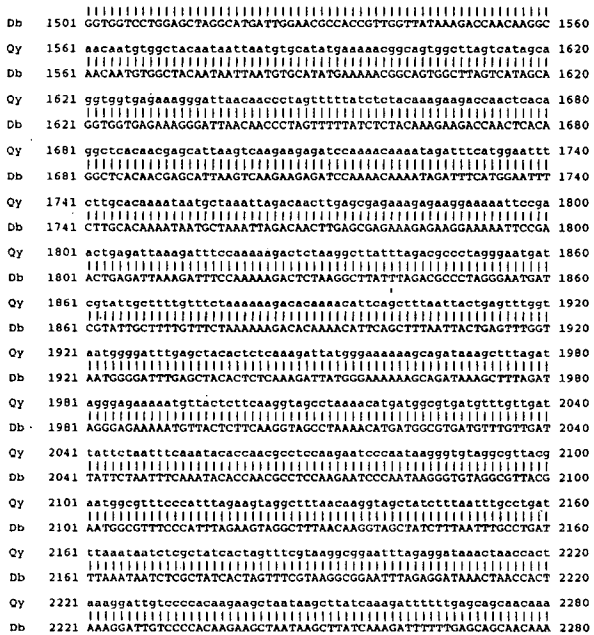


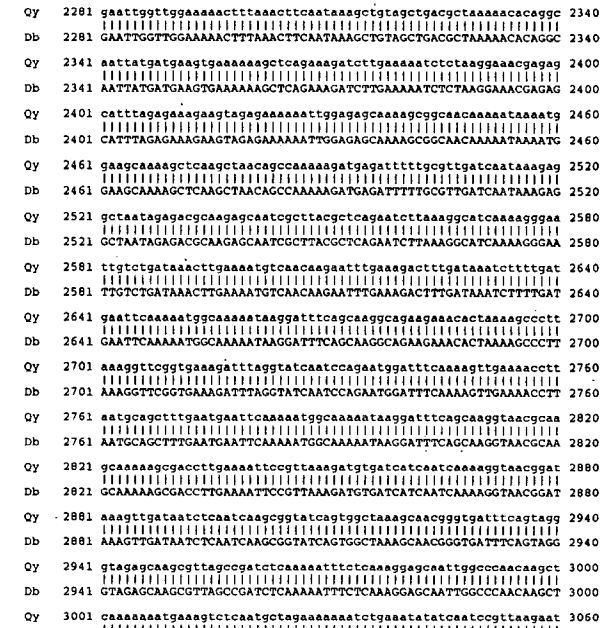
```
     source          1. .5925
                     /organism="Helicobacter pylori"
                     /strain="CCUG 17874"
                     /db_xref="taxon:210"
                     /clone_lib="lambda gt11 ART2/plasmid pHINDIII"
                     /clone="57/D,64/4, 007, 24, A1, G5, A17, B1"
     promoter        1. .534
     prim_transcript 324. .4032
                     /note="transcript B"
     prim_transcript 430. .4032
                     /note="transcript A"
     RBS             525. .529
     gene            535. .3978
                     /gene="cai"
     CDS             535. .3978
                     /gene="cai"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxicity associated immunodominant antigen"
                     /protein_id="CAA49633.1"
                     /db_xref="GI:394913"
                     /db_xref="SWISS-PROT:P80200"
                     /translation="MTNETIDQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDRNDRDNRQAFEGISQLREEYSNKAIKNPTKKNQYFSDPINKSNDLINKDNLIDV
                     ESSTKSPQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPILDDKEKAEPL
                     KSAKQSFAGIIIGNQIRTDQKFMGVFDESLKERQEAEKNGEPTGGDWLDIPLSFIPDK
                     KQSSDVKEAINQEPVPHVQPDIATTTTDIQGLPPEARDLLDERGNPSKPTLGDMEMLD
                     VEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAT
                     VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQLTGSQRALSQEE
                     IQNKIDFMEPLAQNNAKLDNLSEKEKEKFRTEIKDPQKDSKAYLDALGNDRIAFVSKK
                     DTKHSALITEFGNGDLSYTLKDYGKKADKALDREKNVTLQGSLKHDGVMFVDYSNFKY
                     TNASKNPNKGVGVTNGVSHLEVGFNKVAIFNLPDLNNLAITSFVRRNLEDKLTTKGLS
                     PQEANKLIKDFLSSNKELVGKTLNFNKAVADAKNTGNYDEVKKAQKDLEKSLRKREHL
                     EKEVEKKLESKSGNKNKMEAKAQANSQKDEIPALINKEAANRDARAIAYAQNLKGIKRE
                     LSDKLENVNKNLKDFDKSFDEPKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVE
                     NLNAALNEPFKNGKNKDFSKVTQAKSDLENSVKDVIINQKVTDKVDNLNQAVSVAKATG
                     DFSRVEQALADLKNPSKEQLAQQAQKNESLNARKKSEIYQSVKNGVNGTLVGNGLSQA
                     EATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKKKAGQAASLEEPIYAQ
                     VAKKVNAKIDRLNQIASGLGVVGQAAGFPLKRHDKVDDLSKVGLSRNQELAQKIDNLN
                     QAVSEAKAGFPGNLEQTIDKLKDSTKHNPMNLWVESAKKVPASLSAKLDNYATNSHIR
                     INSNIKNGAINEKATGMLTQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMKD
                     YSDSPKFSTKLNNAVKDTNSGFTQFLTNAFSTASYYCLARENAEHGIKNVNTKGGPQK
                     S"
     terminator      4010. .4032
BASE COUNT     2076 a    1057 c    1138 g    1654 t
ORIGIN


  Query Match           100.0%;  Score 5925;  DB 2;  Length 5925;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 5925;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```









```
Qy   3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 GGTGTGAATGGAACCCTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACTCTTTCT 3120

Qy   3121 aaaaactttccggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAAAACTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAAT 3180

Qy   3181 aacaataatggactcaaaaacgaaccccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AACAATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGG 3240

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 CAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAA 3300

Qy   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 ATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC 3360

Qy   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAA 3420

Qy   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTT 3480

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaaacacaatcccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATG 3540

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 AATCTATGGGTTGAAAGTGCAAAAAAAGTACCTGCTAGTTTGTCAGCGAAACTAGACAAT 3600

Qy   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 TACGCTACTAACAGCCACATACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAA 3660

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaacccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 AAAGCGACCGGCATGCTAACGCAAAAAAACCCTGAGTGGCTCAAGCTCGTGAATGATAAG 3720

Qy   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 ATAGTTGCGCATAATGTAGGAAGCGTTCCTTTGTCAGAGTATGATAAAATTGGCTTCAAC 3780

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   3781 CAGAAGAATATGAAAGATTATTCTGATTCGTTCAAGTTTTCCACCAAGTTGAACAATGCT 3840

Qy   3841 gtaaaagacactaatttctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 GTAAAAGACACTAATTCTGGCTTTACGCAATTTTTAACCAATGCATTTTCTACAGCATCT 3900

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TATTACTGCTTGGCGAGAGAAAATGCGGAGCATGGAATCAAGAACGTTAATACAAAAGGT 3960

Qy   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GGTTTCCAAAAATCTTAAAGGATTAAGGAATACCAAAAACGCAAAAACCACCCCTTGCTA 4020

Qy   4021 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttgggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AAAGCGAGGGGTTTTTAATACTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGA 4080

Qy   4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 TGAATGCTACCAATTCATGGTATCATATCCCCATACATTCGTATCTAGCGTAGGAAGTGT 4140

Qy   4141 gcaaagttacgccttttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 4200

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 AACTCTGTAAAATCCCTATTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 4260

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 4320

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAA 4380

Qy   4381 aagacttaaccctttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 AAGACTTAACCCTTTGCTTAAAATTAAGTTTGATTGTGCTAGTGGGTTCGTGCTATAGTG 4440

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 CGAAAATTAATTAAGGGTTATAAAGAGAGCATAAACTAGAAAAAACAAGTAGCTATAACA 4500

Qy   4501 aagatcaagttcaaaaaatcatagagcttttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 AAGATCAAGTTCAAAAAATCATAGAGCTTTTAGAGCAAATTGATCGCGCTCTTAACCAAA 4560

Qy   4561 gaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaaagacgcttgc 4620
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 GAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAAGACGCTTGC 4620

Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 GATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACTGTGG 4680

Qy   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 ATGCCTTTCAAGGTGAAGAGGCAGATATTATTATTTATTCCACCGTGAAAACTTGTGGTA 4740

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 ATCTTTCTTTCTTGCTAGATTCTAAACGCTTGAATGTGGCTATTTCTAGGGCAAAAGAAA 4800

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 ATCTCATTTTTGTGGGTAAAAAGTCTTTCTTTGAGAATTTATGAAGCGATGAGAAGAATA 4860

Qy   4861 tctttagcgctattttgcaagtctgtagatagggtaatctttttccaaagataatcattaga 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 TCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 4920

Qy   4921 cattcttcgcttcaaaacgctttcataaaatctctctaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACACAAT 4980

Qy   4981 acccttatagtgtgagctatagccccttttttgggaattgagttattttgacttttaaattt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 ACCCTTATAGTGTGAGCTATAGCCCCTTTTTGGGAATTGAGTTATTTTGACTTTAAATTT 5040

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCC 5100

Qy   5101 gctcgcatgaaattccacctttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 GCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATTTTTTTTTAAGGGCGTATTTTTG 5160

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggggtgttgaaagggc 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 CTGCAAATATCCTACAATAGCATCGCCCGAATGGATGAGTAGGGGGGGGTGTTGAAAGGGC 5220

Qy   5221 aaaatgctccatcaaaatagccctcaatttttttgagcgattaagggaaaatgcgtgcaacc 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 AAAATGCTCCATAAAATAGCCCTCAATTTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 5280

Qy   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagttttct 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 TAAAATAATCACTTCGGGAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTTTCT 5340
```



```
Qy   5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 AACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAAGAGAAGTGGCTAAA 5400

Qy   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 TGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGGATC 5460

Qy   5461 gtcgcttttgtccctagcactaaaatagggggcgtttttatctttacttgtcgcttgatc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 GTCGCTTTTGTCCCTAGCACTAAAATAGGGGCGTTTTTATCTTTACTTGTCGCTTGATC 5520

Qy   5521 gctaaaatgcttggctcaatcacgcccacaatagggattttggaatgcttttgcatctct 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 GCTAAAATGCTTGGCTCAATCACGCCCACAATAGGGATTTTGGAATGCTTTTGCATCTCT 5580

Qy   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 TCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGGT 5640

Qy   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 TTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTGCCA 5700

Qy   5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 TAAGGCACTCTAGCCGTATCGCCATAATAGATGATTTCATCAAATAATTGCGCTTTTAAA 5760

Qy   5761 aggctttttaaaacgctaaacccctcccacaccgctatcaaaaacgcctattttcatgaca 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 AGGCTTTTTAAAACGCTAAACCCTCCCACACCGCTATCAAAAACGCCTATTTTCATGACA 5820

Qy   5821 ctttttttaatttaatgggattaattagggatttttatttttcattcattaagtttaaaaat 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 CTTTTTTAATTTAATGGGATTAATTAGGGATTTTATTTTTCATTCATTAAGTTTAAAAAT 5880

Qy   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
          |||||||||||||||||||||||||||||||||||||||||||||
Db   5881 TCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATAGACAAAGCTT 5925
```

```
RESULT   3
A76409
LOCUS       A76409       5925 bp    DNA             PAT       19-OCT-1999
DEFINITION  Sequence 4 from Patent WO9318150.
ACCESSION   A76409
VERSION     A76409.1  GI:6088449
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
```

```
REFERENCE   1  (bases 1 to 5925)
  AUTHORS   Covacci,A. and Bugnoli,M.
  TITLE     HELICOBACTER PYLORI PROTEINS USEFUL FOR VACCINES AND DIAGNOSTICS
  JOURNAL   Patent: WO 9318150-A 16-SEP-1993;
            SCLAVO BIOCINE SPA (IT)
FEATURES             Location/Qualifiers
     source          1. .5925
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     2076 a    1057 c    1138 g    1654 t
ORIGIN
```

Query Match 100.0%; Score 5925; DB 81; Length 5925;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5925; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ctccatttttaagcaactccatagaccactaaagaaaacttttttttgaggctatctttgaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCCATTTTAAGCAACTCCATAGACCACTAAAGAAACTTTTTTTGAGGCTATCTTTGAAA 60

Qy      61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatccacaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCTGTCCTATTGATTTGTTTTCCATTTTGTTTCCCATGTGGATCTTGTGGATCACAAAC 120

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTTAATTATACATGCTATAGTAAGCATGACACACAAACCAAACTATTTTTAGAACGCTT 180

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CATGTGCTCACCTTGACTAACCATTTCTCCAACCATACTTTAGCGTTGCATTTGATTTCT 240

Qy     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttttagcaaattt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCAAAAAGATTCATTTCTTATTTCTTGTTCTTATTAAAGTTCTTTCATTTTAGCAAATTT 300

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTGTTAATTGTGGGTAAAAATGTGAATCGTCCTAGCCTTTAGACGCCTGCAACGATCGGG 360

Qy     361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaatgcttgatattgttgtgtat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTTTTTTCAATATTAATAATGATTAATGAAAAAAAAAAAAATGCTTGATATTGTTGTAT 420

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AATGAGAATGTTCAAAGACATGAATTGACTACTCAAGCGTGTAGCGATTTTTAGCAGTCT 480

Qy     481 ttgacactaacaagataccgatagtatgaaactaggtatagtaaggagaaacaatgact 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db     481 TTGACACTAACAAGATACCGATAGGTATGAAACTAGGTATAGTAAGGAGAAACAATGACT 540

Qy     541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AACGAAACCATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATC 600

Qy     601 aataatcttcaagtagctttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGAT 660

Qy     661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAAGGAATCTCG 720

Qy     721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTAT 780

Qy     781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTA 840

Qy     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatttttcacaagt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGT 900

Qy     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATG 960

Qy     961 gaaaatatcatacaaccccctatccttgatgataaagagaaagcggagtttttgaaatct 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GAAAATATCATACAACCCCCTATCCTTGATGATAAAGAGAAAGCGGAGTTTTTGAAATCT 1020

Qy    1021 gccaaaacaatcttttgcaggaatcattatagggaatccaaatccgaacggatcaaaagttc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCCAAACAATCTTTTGCAGGAATCATTATAGGGAATCAAATCCGAACGGATCAAAAGTTC 1080

Qy    1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaaatggagagcct 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ATGGGCGTGTTTGATGAGTCCTTGAAAGAAAGGCAAGAAGCAGAAAAAAATGGAGAGCCT 1140

Qy    1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaaacaatcttct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ACTGGTGGGGATTGGTTGGATATTTTTCTCTCATTTATATTTGACAAAAAACAATCTTCT 1200

Qy    1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GATGTCAAAGAAGCAATCAATCAAGAACCAGTTCCCCATGTCCAACCAGATATAGCCACT 1260

Qy    1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ACCACCACCGACATACAAGGCTTACCGCCTGAAGCTAGAGATTTACTTGATGAAAGGGGT 1320

Qy    1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 AATTTTTCTAAATTCACTCTTGGCGATATGGAAATGTTAGATGTTGAGGGAGTCGCTGAC 1380

Qy    1381 attgatcccaaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 ATTGATCCCAATTACAAGTTCAATCAATTATTGATTCACAATAACGCTCTGTCTTCTGTG 1440

Qy    1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TTAATGGGGAGTCATAATGGCATAGAACCTGAAAAAGTTTCATTGTTGTATGGGGGCAAT 1500

Qy    1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GGTGGTCCTGGAGCTAGGCATGATTGGAACGCCACCGTTGGTTATAAAGACCAACAAGGC 1560

Qy    1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 1620

Qy    1621 ggtggtgagaaagggattaacaaccctagttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 1680

Qy    1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaatttt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 1740

Qy    1741 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCGAGAAAGAGAAGGAAAAATTCCGA 1800

Qy    1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 ACTGAGATTAAAGATTTCCAAAAAGACTCTAAGGCTTATTTAGACGCCCTAGGGAATGAT 1860

Qy    1861 cgtattgctttttgtttctaaaaaaagacacaaaacattcagctttaatttactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 CGTATTGCTTTTTGTTTCTAAAAAAGACACAAAACATTCAGCTTTAATTACTGAGTTTGGT 1920

Qy    1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 AATGGGGATTTGAGCTACACTCTCAAAGATTATGGGAAAAAAGCAGATAAAGCTTTAGAT 1980

Qy    1981 agggagaaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 AGGGAGAAAAATGTTACTCTTCAAGGTAGCCTAAAACATGATGGCGTGATGTTTGTTGAT 2040
```



```
Qy    2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 TATTCTAATTTCAAATACACCAACGCCTCCAAGAATCCCAATAAGGGTGTAGGCGTTACG 2100

Qy    2101 aatgggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 AATGGCGTTTCCCATTTAGAAGTAGGCTTTAACAAGGTAGCTATCTTTAATTTGCCTGAT 2160

Qy    2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 TTAAATAATCTCGCTATCACTAGTTTCGTAAGGCGGAATTTAGAGGATAAACTAACCACT 2220

Qy    2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 AAAGGATTGTCCCCACAAGAAGCTAATAAGCTTATCAAAGATTTTTGAGCAGCAACAAA 2280

Qy    2281 gaattggttggaaaaacttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 GAATTGGTTGGAAAAACTTTAAACTTCAATAAAGCTGTAGCTGACGCTAAAAACACAGGC 2340

Qy    2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 AATTATGATGAAGTGAAAAAAGCTCAGAAAGATCTTGAAAAATCTCTAAGGAAACGAGAG 2400

Qy    2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 CATTTAGAGAAAGAAGTAGAGAAAAAATTGGAGAGCAAAAGCGGCAACAAAAATAAAATG 2460

Qy    2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 GAAGCAAAAGCTCAAGCTAACAGCCAAAAAGATGAGATTTTTGCGTTGATCAATAAAGAG 2520

Qy    2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 GCTAATAGAGACGCAAGAGCAATCGCTTACGCTCAGAATCTTAAAGGCATCAAAAGGGAA 2580

Qy    2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 TTGTCTGATAAACTTGAAAATGTCAACAAGAATTTGAAAGACTTTGATAAATCTTTTGAT 2640

Qy    2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagccctt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 GAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGCAGAAGAAACACTAAAAGCCCTT 2700

Qy    2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaacctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 AAAGGTTCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAAAGTTGAAAACCTT 2760

Qy    2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 AATGCAGCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGTAACGCAA 2820
```

```
Qy   2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GCAAAAAGCGACCTTGAAAATTCCGTTAAAGATGTGATCATCAATCAAAAGGTAACGGAT 2880

Qy   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 AAAGTTGATAATCTCAATCAAGCGGTATCAGTGGCTAAAGCAACGGGTGATTTCAGTAGG 2940

Qy   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaacaagct 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 GTAGAGCAAGCGTTAGCCGATCTCAAAAATTTCTCAAAGGAGCAATTGGCCCAACAAGCT 3000

Qy   3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 CAAAAAAATGAAAGTCTCAATGCTAGAAAAAAATCTGAAATATATCAATCCGTTAAGAAT 3060

Qy   3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 GGTGTGAATGGAACCCTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACTCTTTCT 3120

Qy   3121 aaaaactttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAAAACTTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAAT 3180

Qy   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AACAATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGG 3240

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 CAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAA 3300

Qy   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 ATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC 3360

Qy   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAA 3420

Qy   3421 gaattggctcagaaaatttgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTT 3480

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATG 3540

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   3541 AATCTATGGGTTGAAAGTGCAAAAAAAGTACCTGCTAGTTTGTCAGCGAAACTAGACAAT 3600

Qy   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 TACGCTACTAACAGCCACATACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAA 3660

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 AAAGCGACCGGCATGCTAACGCAAAAAAACCCTGAGTGGCTCAAGCTCGTGAATGATAAG 3720

Qy   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 ATAGTTGCGCATAATGTAGGAAGCGTTCCTTTGTCAGAGTATGATAAAATTGGCTTCAAC 3780

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 CAGAAGAATATGAAAGATTATTCTGATTCGTTCAAGTTTTCCACCAAGTTGAACAATGCT 3840

Qy   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcattttctacagcatct 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 GTAAAAGACACTAATTCTGGCTTTACGCAATTTTTAACCAATGCATTTTCTACAGCATCT 3900

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TATTACTGCTTGGCGAGAGAAAATGCGGAGCATGGAATCAAGAACGTTAATACAAAAGGT 3960

Qy   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GGTTTCCAAAAATCTTAAAGGATTAAGGAATACCAAAAACGCAAAAACCACCCCTTGCTA 4020

Qy   4021 aaagcgagggggtttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AAAGCGAGGGGTTTTTTAATACTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGA 4080

Qy   4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 TGAATGCTACCAATTCATGGTATCATATCCCCATACATTCGTATCTAGCGTAGGAAGTGT 4140

Qy   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 4200

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 AACTCTGTAAAATCCCTATTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 4260

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcacctaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 4320

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAA 4380

Qy   4381 aagacttaacccttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 AAGACTTAACCCTTGCTTAAAATTAAGTTTGATTGTGCTAGTGGGTTCGTGCTATAGTG 4440

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 CGAAAATTAATTAAGGGTTATAAAGAGAGCATAAACTAGAAAAAACAAGTAGCTATAACA 4500

Qy   4501 aagatcaagttcaaaaaatcatagagctttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 AAGATCAAGTTCAAAAAATCATAGAGCTTTAGAGCAAATTGATCGCGCTCTTAACCAAA 4560

Qy   4561 gaaaaatcagaaaaaccataggaattatcacacctataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 GAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAAGACGCTTGC 4620

Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacacgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 GATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACGTGG 4680

Qy   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 ATGCCTTTCAAGGTGAAGAGGCAGATATTATTATTTATTCCACCGTGAAAACTTGTGGTA 4740

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 ATCTTTCTTTCTTGCTAGATTCTAAACGCTTGAATGTGGCTATTTCTAGGGCAAAAGAAA 4800

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 ATCTCATTTTTGTGGGTAAAAAGTCTTTCTTTGAGAATTTATGAAGCGATGAGAAGAATA 4860

Qy   4861 tctttagcgctattttgcaagtctgtagataggtaatctttttccaaagataatcattaga 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 TCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 4920

Qy   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACACAAT 4980

Qy   4981 acccttatagtgtgagctatagccccctttttgggaatttgagttattttgactttaaatttt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 ACCCTTATAGTGTGAGCTATAGCCCCTTTTTGGGAATTGAGTTATTTTGACTTTAAATTT 5040

Qy   5041 ttattagcgttacaatttgagccattctttagcttgttttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCC 5100
```



```
Qy   5101 gctcgcatgaaattccacttttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 GCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATTTTTTTTAAGGGCGTATTTTTG 5160

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggtgttgaaaggggc 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 CTGCAAATATCCTACAATAGCATCGCCCGAATGGATGAGTAGGGGGGGTGTTGAAAGGGC 5220

Qy   5221 aaaatgctccataaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 AAAATGCTCCATAAAATAGCCCTCAATTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 5280

Qy   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 TAAAATAATCACTTCGGGAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTTTCT 5340

Qy   5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 AACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAAGAGAAGTGGCTAAA 5400

Qy   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 TGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGGATC 5460

Qy   5461 gtcgcttttgtccctagcactaaaatagggcgtttttatcttttacttgtcgcttgatc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 GTCGCTTTTGTCCCTAGCACTAAAATAGGGGCGTTTTTATCTTTTACTTGTCGCTTGATC 5520

Qy   5521 gctaaaatgcttggctcaatcacgcccacaataggggattttggaatgcttttgcatctct 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 GCTAAAATGCTTGGCTCAATCACGCCCACAATAGGGGATTTTGGAATGCTTTTGCATCTCT 5580

Qy   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 TCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGGT 5640

Qy   5641 ttgaaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 TTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTGCCA 5700

Qy   5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgctttttaaa 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 TAAGGCACTCTAGCCGTATCGCCATAATAGATGATTTCATCAAATAATTGCGCTTTTTAAA 5760

Qy   5761 aggctttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctatttttcatgaca 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 AGGCTTTTTAAAACGCTAAACCCTCCCACACCGCTATCAAAAACGCCTATTTTCATGACA 5820

Qy   5821 ctttttaatttaatgggattaattagggatttttatttttcattcattaagtttaaaaaat 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 CTTTTTAATTTAATGGGATTAATTAGGGATTTTATTTTTCATTCATTAAGTTTAAAAAT 5880
```

```
Qy   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
          |||||||||||||||||||||||||||||||||||||||||||||
Db   5881 TCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATAGACAAAGCTT 5925


RESULT   4
AF282853
LOCUS       AF282853    20229 bp    DNA             BCT       16-AUG-2000
DEFINITION  Helicobacter pylori cagI pathogenicity island, complete sequence.
ACCESSION   AF282853 U60176
VERSION     AF282853.1  GI:8843984
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 20229)
  AUTHORS   Censini,S., Lange,C., Xiang,Z., Crabtree,J.E., Ghiara,P.,
            Borodovsky,M., Rappuoli,R. and Covacci,A.
  TITLE     cag, a pathogenicity island of Helicobacter pylori, encodes type
            I-specific and disease-associated virulence factors
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (25), 14648-14653 (1996)
  MEDLINE   97121442
   PUBMED   8962108
REFERENCE   2  (bases 1 to 20229)
  AUTHORS   Covacci,A. and Rappuoli,R.
  TITLE     Tyrosine-phosphorylated bacterial proteins: Trojan horses for the
            host cell
  JOURNAL   J. Exp. Med. 191 (4), 587-592 (2000)
  MEDLINE   20150112
   PUBMED   10684850
REFERENCE   3  (bases 1 to 20229)
  AUTHORS   Censini,S., Rappuoli,R., Lange,C. and Covacci,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-1996) Molecular Biology, Chiron-Biocine, Via
            Fiorentina 1, Siena, SI 53100, Italy
REFERENCE   4  (bases 1 to 20229)
  AUTHORS   Censini,S., Rappuoli,R. and Covacci,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) Molecular Biology, Chiron-Biocine, Via
            Fiorentina 1, Siena, SI 53100, Italy
  REMARK    Sequence update by submitter
COMMENT     On Jun 30, 2000 this sequence version replaced gi:1752680.
FEATURES             Location/Qualifiers
     source          1. .20229
                     /organism="Helicobacter pylori"
                     /strain="CCUG 17874; NCTC 11638"
                     /db_xref="taxon:210"
                     /note="type I-specific and disease-associated virulence
                     factor genes; partial sequence of cag pathogenicity island
```



```
                     presented in AF282852"
     source          1. .1880
                     /organism="Helicobacter pylori"
                     /insertion_seq="IS605"
                     /strain="CCUG 17874; NCTC 11638"
                     /db_xref="taxon:210"
                     /note="GenBank Accession Number U60177"
     gene            complement(83. .522)
                     /gene="tnpA"
     CDS             complement(83. .511)
                     /gene="tnpA"
                     /note="similar to IS200 transposase encoded by GenBank
                     Accession Number U60177"
                     /codon_start=1
                     /transl_table=11
                     /product="TnpA"
                     /protein_id="AAC44689.1"
                     /db_xref="GI:1752683"
                     /translation="MRKNHYPLRGYISTNRSKHNLKAHLILVCKYRKKLLQGDLNNFI
                     KSVIDEIATQSHFIIIAMESDIDHLHLMVQYIPRMSISSIISKIKQITTYRVWHDKRF
                     IPLLQKHFWKEKTFWTDGFFVCSIGQANPETIKAYIENQG"
     RBS             complement(519. .522)
                     /gene="tnpA"
                     /note="putative"
     gene            570. .1864
                     /gene="tnpB"
     RBS             570. .574
                     /gene="tnpB"
                     /note="putative"
     CDS             581. .1864
                     /gene="tnpB"
                     /note="similar to IS1341 transposase encoded by GenBank
                     Accession Number U60177"
                     /codon_start=1
                     /transl_table=11
                     /product="TnpB"
                     /protein_id="AAF80205.1"
                     /db_xref="GI:8843985"
                     /translation="MLNAIKFRIYPNAQQKELISKHFGCSRVVYNYFLDYRQKQYAKG
                     IKETYFTMQKVLTQIKHQEKYHYLNECNSQSLQMALRQLVSAYDNFFSKRARYPKFKS
                     KKNAKQSFAIPQNIEIKTETQTIALPKFKEGIRAKLWRELPKDSVIKQAFISCIADQY
                     FCSISYETKEPIPKPTIIKKAVGLDMGLRTLIVTSDKIEYPHIRFYQKLEKKLTKAQR
                     RLSKKVKGSNNRKKQAKKVARLHLACSNTRDDYLHKISNEITNQYDLIGVETLNVKGL
                     MRTYHSKSLANASWGKFLTMLKYKAQRKAKTLLGIDRFFPSSQLCSYCGFNTGKKHEN
                     ITKFTCPHCNITHHRDYNASVNIRNYALGMLDDRHKIKIDKSRVGIIRTDYAHYTDER
                     IKACGASSNGVISKYGNILDLASYGAMKQEKAQSL"
     misc_feature    1881. .19352
                     /note="cagI pathogenicity island"
     gene            complement(2065. .2462)
                     /gene="cagQ"
     CDS             complement(2065. .2445)
```



```
                     /gene="cagQ"
                     /note="HP535; cag14; JHP483"
                     /codon_start=1
                     /transl_table=11
                     /product="CagQ"
                     /protein_id="AAC44691.1"
                     /db_xref="GI:1752685"
                     /translation="MLPTKTRIRDPNKQELTQPKIKGLIMGKILASLLGGGTNLFTGL
                     SSDLFSMILNFLFFLMLMMGLNEALGKKFNLPMDNIKNFMAEVLKNGFDSIKNMGSAL
                     VGNGFGSSKSDKTANKMSVSQVRL"
     RBS             2143. .2146
                     /gene="cagR"
                     /note="putative"
     gene            2143. .2455
                     /gene="cagR"
     CDS             2153. .2455
                     /gene="cagR"
                     /codon_start=1
                     /transl_table=11
                     /product="CagR"
                     /protein_id="AAC44692.1"
                     /db_xref="GI:1752686"
                     /translation="MFLILSNPFFSTSAIKFLILSIGKLNFFPNASLSPIINIRKNKK
                     FNIIENKSLDKPVKRFVPPPNKEAKIFPMISPFIFGCVSSCLPGSLMRVLVGSISQ"
     RBS             complement(2459. .2462)
                     /gene="cagP"
                     /note="putative"
     gene            complement(2878. .3311)
                     /gene="cagP"
     CDS             complement(2878. .3210)
                     /gene="cagP"
                     /note="HP536; cag1; JHP484"
                     /codon_start=1
                     /transl_table=11
                     /product="CagP"
                     /protein_id="AAF80206.1"
                     /db_xref="GI:8843986"
                     /translation="MSKLKQNFLQFKYSFNKHLDKYSLYYRLFNISSIVIGFLIGLFS
                     YGAGVILVYPILPLFALIIRPSFFYYTTYLLLLVSLSIISKYYLLSHAKFTMKLIILM
                     TQWQNWFL"
     RBS             complement(3219. .3221)
                     /gene="cagP"
                     /note="putative"
     gene            complement(3306. .3536)
                     /gene="cagO"
     CDS             complement(3306. .3521)
                     /gene="cagO"
                     /codon_start=1
                     /transl_table=11
                     /product="CagO"
                     /protein_id="AAC44694.1"
```



```
                     /db_xref="GI:1752688"
                     /translation="MFFLLVCGLLVFFKFLLRLFLYNRFVFFRWKTPLFFNRCFLFFV
                     WHKQTNRWFVLYHMLVSASGVFFEIWS"
     RBS             complement(3533. .3536)
                     /gene="cagO"
                     /note="putative"
     RBS             3628. .3631
                     /gene="cagM"
                     /note="putative"
     gene            3628. .4777
                     /gene="cagM"
     CDS             3647. .4777
                     /gene="cagM"
                     /note="HP537; cag16; JHP485"
                     /codon_start=1
                     /transl_table=11
                     /product="CagM"
                     /protein_id="AAC44695.1"
                     /db_xref="GI:1752689"
                     /translation="MLAKIVFSSLVAFGVLSANVEQFGSFFNEIKKEQEEVAAKEDAL
                     KARKKLLNNTHDFLEDLIFRKQKIKELMDHRAKVLSDLENKYKKEKEALEKETRGKIL
                     TAKSKAYGDLEQALKDNPLYKKLLPNPYAYVLNQETFTKEDKERLSYYYPQVKTSSIF
                     KKTTATTKDKAQALLQMGVFSLDEEQNKKASRLALSYKQAIEEYSNNISNLLSRKELD
                     NIDYYLQLERNKFDSKAKDIAQKATNTLIFNSERLAFSMAIDKINEKYLKGYEAFSNL
                     LKNVKDDVELNTLTKNFTNQKLSPAQKQKLCLLVLDSFNPDTQSKKSILKKTNEYNIF
                     VDSDPMMSDKTTMQKEHYKIFNFFKTVVSAYRNNVAKNNPPE"
     RBS             4781. .4785
                     /gene="cagN"
                     /note="putative"
     gene            4781. .5712
                     /gene="cagN"
     CDS             4792. .5712
                     /gene="cagN"
                     /note="HP538; cag17; JHP486"
                     /codon_start=1
                     /transl_table=11
                     /product="CagN"
                     /protein_id="AAF80207.2"
                     /db_xref="GI:9832971"
                     /translation="MKSIFKKLGSVALYSLVVYGGLNAINTALLPSEYKELVALGFKK
                     IKTLYQRHDDKEITKEEKEFATNALREKLRNDRARAEQIQKNIEAFEKKNNSSVQKKA
                     AKHKGLQELNEINANPLNDNPNGNSSTETKSNKDDNFDEMINKVNESFVKPAAPLVPD
                     EWRTPEIEIIINECIISSNDYDGLRKCLIKDIKDQKILAPLLEKIQEIETENNKFSRQ
                     HLSGLKLTLNNSNNRTFLIASCAICEKRKKEMEQENNYQDTTNASEFGTTDTKENEAK
                     DTAFSNNRSKSELPNSVINQIEQSIAHGKK"
     gene            complement(5794. .6507)
                     /gene="cagL"
     CDS             complement(5794. .6507)
                     /gene="cagL"
                     /note="HP539; cag18; JHP487"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="CagL"
                     /protein_id="AAC44697.1"
                     /db_xref="GI:1752691"

  Query Match           95.2%;  Score 5642;  DB 2;  Length 20229;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 5912;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy       1 ctccatttttaagcaactccatagaccactaaagaaactttttttgaggctatctttgaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14304 CTCCATTTTAAGCAACTCCATAGACCACTAAAGAAACTTTTTTTGAGGCTATCTTTGAAA 14363

Qy      61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14364 ATCTGTCCTATTGATTTGTTTTCCATTTTGTTTCCCATGTGGATCTTGTGGATCACAAAC 14423

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaaacgctt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14424 GCTTAATTATACATGCTATAGTAAGCATGACACACAAACCAAACTATTTTTAGAAACGCTT 14483

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgattttct 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14484 CATGTGCTCACCTTGACTAACCATTTCTCCAACCATACTTTAGCGTTGCATTTGATTTCT 14543

Qy     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttagcaaatttt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14544 TCAAAAAGATTCATTTCTTATTTCTTGTTCTTATTAAAGTTCTTTCATTTTAGCAAATTT 14603

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagccttttagacgcctgcaacgatcggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14604 TTGTTAATTGTGGGTAAAAATGTGAATCGTCCTAGCCTTTAGACGCCTGCAACGATCGGG 14663

Qy     361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaatgcttgatattgttgtat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14664 CTTTTTTCAATATTAATAATGATTAATGAAAAAAAAAAAAAATGCTTGATATTGTTGTAT 14723

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14724 AATGAGAATGTTCAAAGACATGAATTGACTACTCAAGCGTGTAGCGATTTTTAGCAGTCT 14783

Qy     481 ttgacactaacaagatacgatgaggtatgaaactaggtatagtaaggagaaacaatgact 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14784 TTGACACTAACAAGATACCGATAGGTATGAAACTAGGTATAGTAAGGAGAAACAATGACT 14843

Qy     541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14844 AACGAAACCATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATC 14903

Qy     601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   14904 AATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGAT 14963

Qy     661 caaaaaccaatcgttgataagaacgataggggataacaggcaagcttttgaaggaatctcg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14964 CAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAAGGAATCTCG 15023

Qy     721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15024 CAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTAT 15083

Qy     781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15084 TTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTA 15143

Qy     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15144 GAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGT 15203

Qy     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15204 TGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATG 15263

Qy     961 gaaaatatcatacaaccccctatccttgatgataaagagaaagcggagtttttgaaatct 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15264 GAAAATATCATACAACCCCCTATCCTTGATGATAAAGAGAAAGCGGAGTTTTTGAAATCT 15323

Qy    1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15324 GCCAAACAATCTTTTGCAGGAATCATTATAGGGAATCAAATCCGAACGGATCAAAAGTTC 15383

Qy    1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15384 ATGGGCGTGTTTGATGAGTCCTTGAAAGAAAGGCAAGAAGCAGAAAAAAATGGAGAGCCT 15443

Qy    1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15444 ACTGGTGGGGATTGGTTGGATATTTTTCTCTCATTTATATTTGACAAAAAACAATCTTCT 15503

Qy    1201 gatgtcaaagaagcaatcaatcaagaaccagttcccccatgtccaaccagatatagccact 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15504 GATGTCAAAGAAGCAATCAATCAAGAACCAGTTCCCCATGTCCAACCAGATATAGCCACT 15563

Qy    1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15564 ACCACCACCGACATACAAGGCTTACCGCCTGAAGCTAGAGATTTACTTGATGAAAGGGGT 15623

Qy    1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgaggggagtcgctgac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15624 AATTTTTCTAAATTCACTCTTGGCGATATGGAAATGTTAGATGTTGAGGGAGTCGCTGAC 15683

Qy    1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15684 ATTGATCCCAATTACAAGTTCAATCAATTATTGATTCACAATAACGCTCTGTCTTCTGTG 15743

Qy    1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15744 TTAATGGGGAGTCATAATGGCATAGAACCTGAAAAAGTTTCATTGTTGTATGGGGGCAAT 15803

Qy    1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15804 GGTGGTCCTGGAGCTAGGCATGATTGGAACGCCACCGTTGGTTATAAAGACCAACAAGGC 15863

Qy    1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15864 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 15923

Qy    1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15924 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 15983

Qy    1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15984 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 16043

Qy    1741 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16044 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCGAGAAAGAGAAGGAAAAATTCCGA 16103

Qy    1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgcccctagggaatgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16104 ACTGAGATTAAAGATTTCCAAAAAGACTCTAAGGCTTATTTAGACGCCCTAGGGAATGAT 16163

Qy    1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16164 CGTATTGCTTTTGTTTCTAAAAAAGACACAAAACATTCAGCTTTAATTACTGAGTTTGGT 16223

Qy    1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16224 AATGGGGATTTGAGCTACACTCTCAAAGATTATGGGAAAAAAGCAGATAAAGCTTTAGAT 16283

Qy    1981 agggagaaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16284 AGGGAGAAAAATGTTACTCTTCAAGGTAGCCTAAAACATGATGGCGTGATGTTTGTTGAT 16343

Qy    2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16344 TATTCTAATTTCAAATACACCAACGCCTCCAAGAATCCCAATAAGGGTGTAGGCGTTACG 16403

Qy    2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16404 AATGGCGTTTCCCATTTAGAAGTAGGCTTTAACAAGGTAGCTATCTTTAATTTGCCTGAT 16463
```



```
Qy    2161 ttaaatatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16464 TTAAATAATCTCGCTATCACTAGTTTCGTAAGGCGGAATTTAGAGGATAAACTAACCACT 16523

Qy    2221 aaaggattgtcccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16524 AAAGGATTGTCCCCACAAGAAGCTAATAAGCTTATCAAAGATTTTTGAGCAGCAACAAA 16583

Qy    2281 gaattggttggaaaaacttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16584 GAATTGGTTGGAAAAACTTAAACTTCAATAAAGCTGTAGCTGACGCTAAAAACACAGGC 16643

Qy    2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16644 AATTATGATGAAGTGAAAAAAGCTCAGAAAGATCTTGAAAAATCTCTAAGGAAACGAGAG 16703

Qy    2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16704 CATTTAGAGAAAGAAGTAGAGAAAAAATTGGAGAGCAAAAGCGGCAACAAAAATAAAATG 16763

Qy    2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16764 GAAGCAAAAGCTCAAGCTAACAGCCAAAAAGATGAGATTTTTGCGTTGATCAATAAAGAG 16823

Qy    2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16824 GCTAATAGAGACGCAAGAGCAATCGCTTACGCTCAGAATCTTAAAGGCATCAAAAGGGAA 16883

Qy    2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16884 TTGTCTGATAAACTTGAAAATGTCAACAAGAATTTGAAAGACTTTGATAAATCTTTTGAT 16943

Qy    2641 gaatttcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagcccct 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16944 GAATTTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGCAGAAGAAACACTAAAAGCCCTT 17003

Qy    2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaacctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17004 AAAGGTTCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAAAGTTGAAAACCTT 17063

Qy    2761 aatgcagctttgaatgaattcaaaaatggcaaaaaataaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17064 AATGCAGCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGTAACGCAA 17123

Qy    2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17124 GCAAAAAGCGACCTTGAAAATTCCGTTAAAGATGTGATCATCAATCAAAAGGTAACGGAT 17183

Qy    2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17184 AAAGTTGATAATCTCAATCAAGCGGTATCAGTGGCTAAAGCAACGGGTGATTTCAGTAGG 17243
```

```
Qy   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaacaagct 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17244 GTAGAGCAAGCGTTAGCCGATCTCAAAAATTTCTCAAAGGAGCAATTGGCCCAACAAGCT 17303

Qy   3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17304 CAAAAAAATGAAAGTCTCAATGCTAGAAAAAAATCTGAAATATATCAATCCGTTAAGAAT 17363

Qy   3061 ggtgtgaatggaacccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17364 GGTGTGAATGGAACCCTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACTCTTTCT 17423

Qy   3121 aaaaacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17424 AAAAACTTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAAT 17483

Qy   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17484 AACAATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGG 17543

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17544 CAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAA 17603

Qy   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17604 ATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC 17663

Qy   3361 ccttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17664 CCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAA 17723

Qy   3421 gaatttggctcagaaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17724 GAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTT 17783

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17784 TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATG 17843

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17844 AATCTATGGGTTGAAAGTGCAAAAAAAGTACCTGCTAGTTTGTCAGCGAAACTAGACAAT 17903

Qy   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17904 TACGCTACTAACAGCCACATACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAA 17963

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db  17964 AAAGCGACCGGCATGCTAACGCAAAAAAACCCTGAGTGGCTCAAGCTCGTGAATGATAAG 18023

Qy   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18024 ATAGTTGCGCATAATGTAGGAAGCGTTCCTTTGTCAGAGTATGATAAAATTGGCTTCAAC 18083

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagtttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18084 CAGAAGAATATGAAAGATTATTCTGATTCGTTCAAGTTTTCCACCAAGTTGAACAATGCT 18143

Qy   3841 gtaaaagacactaatttctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18144 GTAAAAGACACTAATTCTGGCTTTACGCAATTTTTAACCAATGCATTTTCTACAGCATCT 18203

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18204 TATTACTGCTTGGCGAGAGAAAATGCGGAGCATGGAATCAAGAACGTTAATACAAAAGGT 18263

Qy   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18264 GGTTTCCAAAAATCTTAAAGGATTAAGGAATACCAAAAACGCAAAAACCACCCCTTGCTA 18323

Qy   4021 aaagcgagggggttttttaatactgcttagcagaaatcccaatcgtctttagtatttgggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18324 AAAGCGAGGGGTTTTTTAATACTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGA 18383

Qy   4081 tgaatgctaccaatttcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18384 TGAATGCTACCAATTCATGGTATCATATCCCCATACATTCGTATCTAGCGTAGGAAGTGT 18443

Qy   4141 gcaaagttacgccttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18444 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 18503

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18504 AACTCTGTAAAATCCCTATTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 18563

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18564 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 18623

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18624 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAA 18683

Qy   4381 aagacttaacccctttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18684 AAGACTTAACCCTTTGCTTAAAATTAAGTTTGATTGTGCTAGTGGGTTCGTGCTATAGTG 18743

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18744 CGAAAATTAATTAAGGGTTATAAAGAGAGCATAAACTAGAAAAAACAAGTAGCTATAACA 18803

Qy   4501 aagatcaagttcaaaaaatcatagagctttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18804 AAGATCAAGTTCAAAAAATCATAGAGCTTTTAGAGCAAATTGATCGCGCTCTTAACCAAA 18863

Qy   4561 gaaaaatcagaaaaaccataggaattatcacacctttataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18864 GAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAAGACGCTTGC 18923

Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18924 GATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACTGTGG 18983

Qy   4681 atgcctttcaaggtgaagaggcagatattattattattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18984 ATGCCTTTCAAGGTGAAGAGGCAGATATTATTATTTATTCCACCGTGAAAACTTGTGGTA 19043

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19044 ATCTTTCTTTCTTGCTAGATTCTAAACGCTTGAATGTGGCTATTTCTAGGGCAAAAGAAA 19103

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19104 ATCTCATTTTTGTGGGTAAAAAGTCTTTCTTTGAGAATTTACGAAGCGATGAGAAGAATA 19163

Qy   4861 tctttagcgctattttgcaagtctgtagataggtaatctttttccaaagataatcattaga 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19164 TCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 19223

Qy   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19224 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACACAAT 19283

Qy   4981 accctatagtgtgagctatagccccttttttgggaattgagttatttgacttttaaattt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19284 ACCCTTATAGTGTGAGCTATAGCCCCTTTTTGGGAATTGAGTTATTTTGACTTTAAATTT 19343

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19344 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCC 19403

Qy   5101 gctcgcatgaaattccacttttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19404 GCTCGCATGAAATTCCACTTTTAGGGAATGCGTGTGCATTTTTTTTTAAGGGCGTATTTTTG 19463

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggggtgttgaaagggc 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19464 CTGCAAATATCCTACAATAGCATCGCCCGAATGGATGAGTAGGGGGGGTGTTGAAAGGGC 19523
```



```
Qy   5221 aaaatgctccataaaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19524 AAAATGCTCCATAAAATAGCCCTCAATTTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 19583

Qy   5281 taaaataatcacttcggg-aaaatctttaagggagtgaaataataacgcatgcaagtttc 5339
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  19584 TAAAATAATCACTTCGGGTAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTTTC 19643

Qy   5340 taacaattcgcccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaa 5399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19644 TAACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAAGAGAAGTGGCTAA 19703

Qy   5400 atgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggat 5459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19704 ATGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGGAT 19763

Qy   5460 cgtcgcttttgtccctagcactaaaataggggcgtttttatctttacttgtcgcttgat 5519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19764 CGTCGCTTTTGTCCCTAGCACTAAAATAGGGGCGTTTTTATCTTTACTTGTCGCTTGAT 19823

Qy   5520 cgctaaaatgcttggctcaatcacgcccacaataggggattttggaatgcttttgcatctc 5579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19824 CGCTAAAATGCTTGGCTCAATCACGCCCACAATAGGGATTTTGGAATGCTTTTGCATCTC 19883

Qy   5580 ttctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcgg 5639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19884 TTCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGG 19943

Qy   5640 tttgaaaaaatccaaagcctctaagccaaatttgcttgatcgtagtggggtctttagtgcc 5699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19944 TTTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTGCC 20003

Qy   5700 ataaggcactctagccgtatcgccataataagatgatttcatcaaataattgcgcttttaa 5759
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  20004 ATAAGGCACTCTAGCGCTATCGCCATAATAGATGATTTCATCAAATAATTGCGCTTTTAA 20063

Qy   5760 aaggcttttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctatttttcatgac 5819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20064 AAGGCTTTTTAAAACGCTAAACCCTCCCACACCGCTATCAAAAACGCCTATTTTCATGAC 20123

Qy   5820 acttttttaatttaatgggattaattagggatttttatttttcattcattaagtttaaaaa 5879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20124 ACTTTTTTAATTTAATGGGATTAATTAGGGATTTTATTTTTCATTCATTAAGTTTAAAAA 20183

Qy   5880 ttcttcattgtccttagtttgttgcattttagaata 5915
          ||||||||||||||||||||||||||||||||||||
Db  20184 TTCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATA 20219

RESULT   5
HPY239688
```

```
LOCUS       HPY239688     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain NCTC11638.
ACCESSION   AJ239688
VERSION     AJ239688.1  GI:4467694
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="NCTC 11638"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37799.1"
                     /db_xref="GI:4467695"
                     /translation="EATPNPQQFINNLQVAFLKVDNAVASYDPDQKPIVDKNDRDNRQ
                     AFEGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTKSFQKFGD
                     QRYRIFTSWVSNQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      153 a     84 c     67 g    109 t
ORIGIN
```

```
  Query Match             6.0%;  Score 353;  DB 2;  Length 413;
  Best Local Similarity   99.8%;  Pred. No. 1.5e-156;
  Matches  403;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    580 tttaacccgcagcaatttatcaataatcttcaagtagcttttcttaaagttgataacgct 639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 TTTAACCCGCAGCAATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCT 69
```



```
Qy    640 gtcgcttcatacgatcctgatcaaaaaccaatcgttgataagaacgatagggataacagg 699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 GTCGCTTCATACGATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGG 129

Qy    700 caagcttttgaaggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaat 759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 CAAGCTTTTGAAGGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAAT 189

Qy    760 cctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgatttaatcaac 819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 CCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAAC 249

Qy    820 aaagacaatctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcag 879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 AAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAG 309

Qy    880 cgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacc 939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 CGTTACCGAATTTTCACAAGTTGGGTGTCCAATCAAAACGATCCGTCTAAAATCAACACC 369

Qy    940 cgatcgatccgaaatttttatggaaaaatatcatacaaccccctat 983
          |||||||||||||||||||||||||||||||||||||||||||||
Db    370 CGATCGATCCGAAATTTTATGGAAAATATCATACAACCCCCTAT 413
```

```
RESULT   6
HPY239685
LOCUS       HPY239685     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain 96-228.
ACCESSION   AJ239685
VERSION     AJ239685.1  GI:4467688
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.;
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
```



```
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="96-228"
                     /db_xref="taxon:210"
                     /country="USA:Alaska"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37796.1"
                     /db_xref="GI:4467689"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVASYDPDQKPIVDKNDRDNRQ
                     AFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTKSFQKFGD
                     QRYRIFTSWVSHQNDPSKINIRSIRNFMGNIIQPP"
BASE COUNT      149 a     84 c     68 g    112 t
ORIGIN
```

```
  Query Match             5.3%;  Score 316;  DB 2;  Length 413;
  Best Local Similarity   99.7%;  Pred. No. 6.1e-139;
  Matches  366;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    571 gaagcggcttttaacccgcagcaatttatcaataatcttcaagtagcttttcttaaagtt 630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGTT 60

Qy    631 gataacgctgtcgcttcatacgatcctgatcaaaaaccaatcgttgataagaacgatagg 690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGG 120

Qy    691 gataacaggcaagcttttgaaggaatctcgcaattaagggaagaatactccaataaagcg 750
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    121 GATAACAGGCAAGCTTTTGATGGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCG 180

Qy    751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 240

Qy    811 ttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaattt 870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTT 300

Qy    871 ggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaa 930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAA 360

Qy    931 atcaaca 937
          |||||||
```



```
Db    361 ATCAACA 367


RESULT   7
HPY239686
LOCUS       HPY239686     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain 5-1.
ACCESSION   AJ239686
VERSION     AJ239686.1  GI:4467690
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="5-1"
                     /db_xref="taxon:210"
                     /country="Lithuania"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37797.1"
                     /db_xref="GI:4467691"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVASYDPDQKPIVDKNDRDNRQ
                     AFNGISQLREEYSNKAIKNPTKNQYFSDFINKSNDLINKDNLIDVESSTKSFQKFGD
                     QRYRIFTSWVSHQNDPSKINTRSIRNFMGNIIQPP"
BASE COUNT      150 a     85 c     67 g    111 t
ORIGIN
```

```
  Query Match             4.9%;  Score 289;  DB 2;  Length 413;
  Best Local Similarity   99.5%;  Pred. No. 4.3e-126;
```

```
  Matches  389;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     571 gaagcggctttttaacccgcagcaatttatcaataatcttcaagtagctttttcttaaagtt 630
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGTT 60

Qy     631 gataacgctgtcgcttcatacgatcctgatcaaaaaccaatcgttgataagaacgatagg 690
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGG 120

Qy     691 gataacaggcaagcttttgaaggaatctcgcaattaagggaagaatactccaataaagcg 750
           |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db     121 GATAACAGGCAAGCTTTTAATGGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCG 180

Qy     751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 240

Qy     811 ttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaattt 870
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTT 300

Qy     871 ggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaa 930
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAA 360

Qy     931 atcaacacccgatcgatccgaaattttatgg 961
           |||||||||||||||||||||||||||||||
Db     361 ATCAACACCCGATCGATCCGAAATTTTATGG 391


RESULT   8
HPY269887
LOCUS       HPY269887     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            US101 from United States.
ACCESSION   AJ269887
VERSION     AJ269887.1  GI:5921351
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
```



```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="US101"
                     /db_xref="taxon:210"
                     /country="USA"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56486.1"
                     /db_xref="GI:5921352"
                     /translation="MTNETINQQPQTEXXFNPQRFINNLQVAFLKVDNAVASYDPDQK
                     PIIDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      168 a     95 c     67 g    115 t      4 others
ORIGIN


  Query Match          4.9%;  Score 288;  DB 2;  Length 449;
  Best Local Similarity   99.5%;  Pred. No. 1.3e-125;
  Matches  388;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     594 atttatcaataatcttcaagtagctttttcttaaagttgataacgctgtcgcttcatacga 653
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 ATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGA 119

Qy     654 tcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaagg 713
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||| ||
Db     120 TCCTGATCAAAAACCAATCATTGATAAGAACGATACGGATAACAGGCAAGCTTTTGATGG 179

Qy     714 aatctcgcaatttaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaa 773
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 AATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAA 239

Qy     774 tcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcat 833
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 TCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCAT 299

Qy     834 tgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatttt 893
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 TGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTT 359

Qy     894 cacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaa 953
```



```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 CACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAA 419

Qy     954 ttttatggaaaatatcatacaaccccctat 983
           |||||||||||||||||||||||||||||||
Db     420 TTTTATGGAAAATATCATACAACCCCCTAT 449


RESULT   9
HPY269894
LOCUS       HPY269894     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            TH8832 from Thailand.
ACCESSION   AJ269894
VERSION     AJ269894.1  GI:5921365
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="TH8832"
                     /db_xref="taxon:210"
                     /country="Thailand"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56456.1"
                     /db_xref="GI:5921366"
                     /translation="MANETINQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDALIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      165 a     97 c     71 g    116 t
```



```
ORIGIN


  Query Match          4.7%;  Score 277;  DB 2;  Length 449;
  Best Local Similarity   99.3%;  Pred. No. 2.2e-120;
  Matches  427;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     554 accaacaaccacaaaccgaagcggctttaacccgcagcaatttatcaataatcttcaag 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 ACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAG 79

Qy     614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 TAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCG 139

Qy     674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     140 TTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGATGGAATCTCGCAATTAAGGGAAG 199

Qy     734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 AATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTA 259

Qy     794 tcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaa 853
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     260 TCAATAAGAGCAATGATTTAATCAACAAAGACGCTCTCATTGATGTAGAATCTTCCACAA 319

Qy     854 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 AGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATC 379

Qy     914 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaaatatcatac 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     380 AAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATAC 439

Qy     974 aaccccctat 983
           ||||||||||
Db     440 AACCCCCTAT 449


RESULT  10
HPY269895
LOCUS       HPY269895     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            NL23 from The Netherlands.
ACCESSION   AJ269895
VERSION     AJ269895.1  GI:5921367
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
```

```
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="NL23"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56492.1"
                     /db_xref="GI:5921368"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDALIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      166 a      98 c      70 g     115 t
ORIGIN


  Query Match             4.7%;  Score 277;  DB 2;  Length 449;
  Best Local Similarity   99.3%;  Pred. No. 2.2e-120;
  Matches  427;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 ACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAG 79

Qy    614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 TAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCG 139

Qy    674 ttgataagaacgataggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    140 TTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGATGGAATCTCGCAATTAAGGGAAG 199
```



```
Qy    734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 AATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTA 259

Qy    794 tcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaa 853
          ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db    260 TCAATAAGAGCAATGATTTAATCAACAAAGACGCTCTCATTGATGTAGAATCTTCCACAA 319

Qy    854 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 AGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATC 379

Qy    914 aaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatggaaaatatcatac 973
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 AAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATAC 439

Qy    974 aacccccctat 983
          ||||||||||
Db    440 AACCCCCTAT 449


RESULT  11
HPY269881
LOCUS       HPY269881     449 bp    DNA             BCT      23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            P017 from Portugal.
ACCESSION   AJ269881
VERSION     AJ269881.1  GI:5921339
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="P017"
                     /db_xref="taxon:210"
                     /country="Portugal"
```



```
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56449.1"
                     /db_xref="GI:5921340"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFLKLDNAVASFDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      164 a      95 c      71 g     119 t
ORIGIN


  Query Match             4.6%;  Score 272;  DB 2;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-118;
  Matches  272;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 GGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAG 237

Qy    772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 AATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTC 297

Qy    832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 ATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATT 357

Qy    892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 TTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGA 417

Qy    952 aattttatggaaaatatcatacaaccccctat 983
          |||||||||||||||||||||||||||||||
Db    418 AATTTTATGGAAAATATCATACAACCCCCTAT 449


RESULT  12
HPY239683
LOCUS       HPY239683     413 bp    DNA             BCT      30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain N6.
ACCESSION   AJ239683
VERSION     AJ239683.1  GI:4467684
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
```



```
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="N6"
                     /db_xref="taxon:210"
                     /country="France"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37794.1"
                     /db_xref="GI:4467685"
                     /db_xref="SPTREMBL:Q9Z5M8"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVASFDPDQKPIVDKNDRDNRQ
                     AFEGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTKSFQKFGD
                     QRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      147 a      85 c      70 g     111 t
ORIGIN


  Query Match             4.4%;  Score 259;  DB 2;  Length 413;
  Best Local Similarity   100.0%;  Pred. No. 8e-112;
  Matches  259;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    725 taagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttt 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    155 TAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTT 214

Qy    785 cagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaat 844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    215 CAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTAGAAT 274

Qy    845 cttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagttggg 904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    275 CTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGG 334

Qy    905 tgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatggaaa 964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    335 TGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAAA 394

Qy    965 atatcatacaaccccctat 983
          ||||||||||||||||||
Db    395 ATATCATACAACCCCCTAT 413


RESULT  13
HPY239721
LOCUS       HPY239721     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain #12.
ACCESSION   AJ239721
VERSION     AJ239721.1  GI:4467760
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Heibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Helicobacter pylori"
                     /strain="#12"
                     /db_xref="taxon:210"
                     /country="Guatemala"
     gene            1. .243
                     /gene="cagA"
     CDS             <1. .>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37813.1"
                     /db_xref="GI:4467761"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
```



```
                     DQLTGSQRALSQEEIQNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       96 a     42 c     49 g     56 t
ORIGIN


  Query Match             4.1%;  Score 243;  DB 2;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-104;
  Matches  243;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1537 gttggttataaagaccaacaaggcaacaatgtggctacaataattaatgtgcatatgaaa 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTTGGTTATAAAGACCAACAAGGCAACAATGTGGCTACAATAATTAATGTGCATATGAAA 60

Qy   1597 aacggcagtggcttagtcatagcaggtggtgagaaagggattaacaaccctagtttttat 1656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AACGGCAGTGGCTTAGTCATAGCAGGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTAT 120

Qy   1657 ctctacaaagaagaccaactcacaggctcacaacgagcattaagtcaagaagagatccaa 1716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAA 180

Qy   1717 aacaaaatagatttcatggaatttcttgcacaaaatgctaaattagacaacttgagc 1776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AACAAAATAGATTTCATGGAATTTCTTGCACAAAATGCTAAATTAGACAACTTGAGC 240

Qy   1777 gag 1779
          |||
Db    241 GAG 243


RESULT  14
HPY239730
LOCUS       HPY239730     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain NCTC11638.
ACCESSION   AJ239730
VERSION     AJ239730.1  GI:4467778
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Helicobacter pylori"
                     /strain="NCTC 11638"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1. .243
                     /gene="cagA"
     CDS             <1. .>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37822.1"
                     /db_xref="GI:4467779"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIQNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       96 a     42 c     49 g     56 t
ORIGIN


  Query Match             4.1%;  Score 243;  DB 2;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-104;
  Matches  243;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1537 gttggttataaagaccaacaaggcaacaatgtggctacaataattaatgtgcatatgaaa 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTTGGTTATAAAGACCAACAAGGCAACAATGTGGCTACAATAATTAATGTGCATATGAAA 60

Qy   1597 aacggcagtggcttagtcatagcaggtggtgagaaagggattaacaaccctagtttttat 1656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AACGGCAGTGGCTTAGTCATAGCAGGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTAT 120

Qy   1657 ctctacaaagaagaccaactcacaggctcacaacgagcattaagtcaagaagagatccaa 1716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAA 180

Qy   1717 aacaaaatagatttcatggaatttcttgcacaaaatgctaaattagacaacttgagc 1776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AACAAAATAGATTTCATGGAATTTCTTGCACAAAATGCTAAATTAGACAACTTGAGC 240

Qy   1777 gag 1779
          |||
Db    241 GAG 243


RESULT  15
AF289431
```



```
LOCUS       AF289431      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR61 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289431
VERSION     AF289431.1. GI:9964216
KEYWORDS
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR61"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09880.1"
                     /db_xref="GI:9964217"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTKSFQKFPGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      110 a     61 c     43 g     77 t
ORIGIN


  Query Match             4.0%;  Score 235;  DB 2;  Length 291;
```

```
  Best Local Similarity   100.0%;  Pred. No. 2.2e-100;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 60

Qy    811 ttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaattt 870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTT 120

Qy    871 ggggatcagcgttaccgaatttttcacaagttgggtgtcccatcaaaacgatccgtctaaa 930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAA 180

Qy    931 atcaacacccgatcgatccgaaattttatggaaaatatcatacaaccccctatcc 985
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 235


RESULT  16
I95673
LOCUS       I95673       3648 bp    DNA             PAT       01-DEC-1998
DEFINITION  Sequence 1 from patent US 5733740.
ACCESSION   I95673
VERSION     I95673.1  GI:3940143
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3648)
  AUTHORS   Cover,T.L., Blaser,M.J., Kleanthous,H. and Tummuru,M.K.R.
  TITLE     Taga gene and methods for detecting predisposition to peptic
            ulceration and gastric carcinoma
  JOURNAL   Patent: US 5733740-A 1 31-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .3648
                     /organism="unknown"
BASE COUNT     1298 a    660 c    692 g    998 t
ORIGIN


  Query Match             4.0%;  Score 235;  DB 82;  Length 3648;
  Best Local Similarity   99.7%;  Pred. No. 1.8e-100;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy    610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy    670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy    730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy    790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  17
I95674
LOCUS       I95674       4821 bp    DNA             PAT       01-DEC-1998
DEFINITION  Sequence 3 from patent US 5733740.
ACCESSION   I95674
VERSION     I95674.1  GI:3940144
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4821)
  AUTHORS   Cover,T.L., Blaser,M.J., Kleanthous,H. and Tummuru,M.K.R.
  TITLE     Taga gene and methods for detecting predisposition to peptic
            ulceration and gastric carcinoma
  JOURNAL   Patent: US 5733740-A 3 31-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .4821
                     /organism="unknown"
BASE COUNT     1721 a    874 c    930 g   1296 t
ORIGIN


  Query Match             4.0%;  Score 235;  DB 82;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 1.8e-100;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy    610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy    670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy    730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy    790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  18
HECMAJANT
LOCUS       HECMAJANT    4826 bp    DNA             BCT       11-JUN-1993
DEFINITION  Helicobacter pylori major antigen gene sequence.
ACCESSION   L11714
VERSION     L11714.1  GI:290950
KEYWORDS    major antigen.
SOURCE      Helicobacter pylori (library: lambda ZAP; ATCC 53726) DNA.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 4826)
  AUTHORS   Tummuru,M.K.R., Cover,T.L. and Blaser,M.J.
  TITLE     Cloning and expression of a high molecular mass major antigen of
            Helicobacter pylori
  JOURNAL   Infect. Immun. 61, 1799-1809 (1993)
  MEDLINE   93239281
FEATURES             Location/Qualifiers
     source          1. .4826
                     /organism="Helicobacter pylori"
                     /db_xref="taxon:210"
                     /tissue_lib="lambda ZAP; ATCC 53726"
BASE COUNT     1722 a    875 c    932 g   1297 t
ORIGIN


  Query Match             4.0%;  Score 235;  DB 2;  Length 4826;
  Best Local Similarity   99.7%;  Pred. No. 1.8e-100;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1148

Qy    610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1208

Qy    670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1268

Qy    730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1328

Qy    790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1329 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1374


RESULT  19
AF247651
LOCUS       AF247651     3540 bp    DNA             BCT       13-APR-2000
DEFINITION  Helicobacter pylori cytotoxin-associated protein CagA (cagA) gene,
            complete cds.
ACCESSION   AF247651
VERSION     AF247651.1  GI:7547015
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 3540)
  AUTHORS   Zhang,J.Z.
  TITLE     cagA gene of Helicobacter pylori (SS1 strain)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3540)
  AUTHORS   Zhou,J.C. and Zhang,J.Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Dept. of Diarrhea Disease, Institute of
            Epidemiology and Microbiology, Chinese Academy of Preventive
            Medicine, P.O.Box 5, Changping, Beijing 102206, China
FEATURES             Location/Qualifiers
     source          1. .3540
                     /organism="Helicobacter pylori"
                     /strain="SS1"
                     /db_xref="taxon:210"
                     /clone="pTSS1"
     gene            1. .3540
                     /gene="cagA"
     CDS             1. .3540
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin-associated protein CagA"
                     /protein_id="AAF63759.1"
                     /db_xref="GI:7547016"
                     /translation="MTNETINQQPQTEAAPNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIIDKNDRDNRQAFEGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDDLIDV
```

```
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEPL
                     KSAKQSFAGIIIGNQIRTDQKFMGVFDESLKERQEAEKNGGPTGGDWLDIFLSFIFDR
                     KQSSDVKEAINQEPVPHVQPDIATTTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLD
                     VEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNAT
                     VGYKDQQGNNVATIINVDMKNGSGLVIAGGEKGINNPSFYLYKEDQLTGSQRALSQEE
                     IRNKIDFMEFLAQNVAKLDNLSEKEKEKFRTEIKDFQKDSKAYLDALGNDRIAFVSKK
                     DPKHSALITEPGKGDLSYTLKDYGKKADKALDREKNVTLQGSLKHDGVMFVDYSNFKY
                     TNASKNPNKGVGVTNGVSHLEAGFSKVAVFNLPDLNNLAITSFVRRNLEDKLATKGLS
                     LQEANKLIKDFLSSNKELVGKALNFNKAVAEAKNTGNYDEVKKAQKDLEKSLRKREHL
                     EKEVEKKLESKSGNKNKMEAKSQANSQKDEIFALINKEANRDARAIAYATNLRGVKRE
                     LSDKLENVNKNLKDFSKSFDEFKNGKNKDFSKAEETLKTLKGSVKDLGINPEWISKVE
                     NLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKITDKVDNLNQAVSVAKATG
                     DFSRVEQALADLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGVNGTLVGNGLSKA
                     EATTLSKNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAQVNKKKAGQVVSPEEPIYAQ
                     VAKKVNAKIDQLNQAASGLGGVGQAGFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLG
                     GPFLKKHAKVEDLSKVGLSREQELTQKIDNLNQAVSEAKAGFFGNLEQTIDNLKDST
                     KHNPMNLWAESAKKVPASLSAKLDNYATNSHTRINSNIQNGAINEKATGMLTQKNPEW
                     LKLVNDKIVVHNVGSVPLSEYDKIGFNQKNMKDYSDSPKFSTKLNNAVKDVRSGFTQF
                     LANAPSTGYYCLAEENAEHGIENVNTRGGFQKS"
BASE COUNT      1304 a    625 c    724 g    887 t
ORIGIN

  Query Match             3.8%;  Score 228;  DB 1;  Length 3540;
  Best Local Similarity   99.1%;  Pred. No. 3.9e-97;
  Matches  428;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy     554 accaacaaccacaaaccgaagcggctttttaacccgcagcaatttatcaataatcttcaag 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 ACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAG 79

Qy     614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 TGGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCA 139

Qy     674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     140 TTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGGGAAG 199

Qy     734 aatactccaatagagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 AATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTA 259

Qy     794 tcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaa 853
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     260 TCAATAAGAGCAATGATTTAATCAACAAAGACGATCTCATTGATGTAGAATCTTCCACAA 319

Qy     854 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 AGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATC 379
```



```
Qy     914 aaaacgatccgtctaaaatcaacacccgatcgatccgaaatttatggaaaaatatcatac 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     380 AAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTATGGAAAATATCATAC 439

Qy     974 aaccccctatcc 985
           ||||||||||||
Db     440 AACCCCCTATCC 451


RESULT  20
AF191014
LOCUS       AF191014      1100 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain Indial20A cag pathogenicity island
            polymorphic right end, type Ib motif.
ACCESSION   AF191014
VERSION     AF191014.1  GI:10567577
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 1100)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Helicobacter pylori"
                     /strain="Indial20A"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="India: Calcutta"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ib motif; similar to
                     Helicobacter pylori strain HUP B5 type Ib motif presented
                     in GenBank Accession Number AF190659; does not have 46 bp
```



```
                     deletion compared to Helicobacter pylori strain NCTC11638"
BASE COUNT      361 a    211 c    206 g    322 t
ORIGIN

  Query Match             3.7%;  Score 222;  DB 1;  Length 1100;
  Best Local Similarity   99.4%;  Pred. No. 3e-94;
  Matches  322;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    4141 gcaaagttacgccttttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 223

Qy    4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     224 AACTCTGTAAAATCCCTACTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 283

Qy    4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 343

Qy    4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAA 403

Qy    4381 aagacttaacccttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     404 AAGACTTAACCCTTTGCTTAAAATTAAATTTGATTGTGCTAGTGGGTTCGTGCTATAGTG 463

Qy    4441 cgaaaattaattaagggttataaa 4464
           |||||||||||||||||||||||
Db     464 CGAAAATTAATTAAGGGTTATAAA 487


RESULT  21
HPY239687
LOCUS       HPY239687     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain 96-232.
ACCESSION   AJ239687
VERSION     AJ239687.1  GI:4467692
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
```



```
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="96-232"
                     /db_xref="taxon:210"
                     /country="USA:Alaska"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37798.1"
                     /db_xref="GI:4467693"
                     /translation="EAAFNPQQPINNLQVAFLKVDNAVASYDPDQKPIVDKNDRDNRQ
                     AFDGISQLREEYSNKAIKNPTKKNQYFSDPINKSNDLINKDNLIDVESSTKSFQKFGD
                     QRYRIPTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      150 a     84 c     67 g    112 t
ORIGIN

  Query Match             3.7%;  Score 221;  DB 2;  Length 413;
  Best Local Similarity   99.6%;  Pred. No. 9.7e-94;
  Matches  271;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142 GGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAG 201

Qy     772 aatcagtatttttcagactttatcaataaagagcaatgatttaatcaacaaagacaatctc 831
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     202 AATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATCTAATCAACAAAGACAATCTC 261

Qy     832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     262 ATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATT 321

Qy     892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 951
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     322 TTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGA 381

Qy     952 aattttatggaaaatatcatacaaccccctat 983
           |||||||||||||||||||||||||||||||
```

```
Db    382 AATTTTATGGAAAATATCATACAACCCCCTAT 413


RESULT  22
HPY239689
LOCUS       HPY239689     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain SS1.
ACCESSION   AJ239689
VERSION     AJ239689.1  GI:4467696
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="SS1"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37800.1"
                     /db_xref="GI:4467697"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVASYDPDQKPIIDKNDRDNRQ
                     AFEGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDDLIDVESSTKSFQKFGD
                     QRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      149 a     85 c     69 g    110 t
ORIGIN


  Query Match             3.7%;  Score 221;  DB 2;  Length 413;
  Best Local Similarity   99.6%;  Pred. No. 9.7e-94;
```



```
  Matches  271;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 GGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAG 201

Qy    772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    202 AATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACGATCTC 261

Qy    832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 ATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATT 321

Qy    892 ttcacaagttggggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 TTCACAAGTTGGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGA 381

Qy    952 aattttatggaaaatatcatacaaccccctat 983
          |||||||||||||||||||||||||||||||
Db    382 AATTTTATGGAAAATATCATACAACCCCCTAT 413


RESULT  23
AF289433
LOCUS       AF289433      576 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR2 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289433
VERSION     AF289433.1  GI:9964220
KEYWORDS
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Helicobacter pylori"
```



```
                     /isolate="CR2"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>576
                     /gene="cagA"
     CDS             <1. .>576
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09882.1"
                     /db_xref="GI:9964221"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKKHGKVDDLSKVGLSASP
                     EPIYATIDDLGGPPPLKRHDKVDDLSKVGLSRNQELAQKIDNLNQAVSEAKAGFFGNL
                     EQTIDKLKDSTKHNPMNLWVESAKKVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      206 a    103 c    125 g    142 t
ORIGIN


  Query Match             3.6%;  Score 212;  DB 2;  Length 576;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-89;
  Matches  212;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3358 ttccctttgaaaaggcatgataaagttgatgatctcagtaaggtagggcctttcaaggaat 3417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    346 TTCCCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAAT 405

Qy   3418 caagaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggt 3477
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    406 CAAGAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGT 465

Qy   3478 ttttttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatccc 3537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    466 TTTTTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCC 525

Qy   3538 atgaatctatgggttgaaagtgcaaaaaaagt 3569
          ||||||||||||||||||||||||||||||||
Db    526 ATGAATCTATGGGTTGAAAGTGCAAAAAAGT 557


RESULT  24
HPY269896
LOCUS       HPY269896     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate NL7
```



```
            from The Netherlands.
ACCESSION   AJ269896
VERSION     AJ269896.1  GI:5921369
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="NL7"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56491.1"
                     /db_xref="GI:5921370"
                     /translation="MTNETINQQPQSEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDALIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      166 a     96 c     71 g    116 t
ORIGIN


  Query Match             3.5%;  Score 210;  DB 2;  Length 449;
  Best Local Similarity   99.0%;  Pred. No. 1.7e-88;
  Matches  410;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    570 cgaagcggcttttaacccgcagcaatttatcaataatcttcaagtagcttttcttaaagt 629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     36 CGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGT 95

Qy    630 tgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcgttgataagaacgatag 689
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96 TGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCGTTGATAAGAACGATAG 155

Qy     690 ggataacaggcaagcttttgaaggaatctcgcaattaagggaagaatactccaataaagc 749
           |||||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     156 GGATAATAGGCAAGCTTTTGATGGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGC 215

Qy     750 gatcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatga 809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 GATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGA 275

Qy     810 tttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaatt 869
           ||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db     276 TTTAATCAACAAAGACGCTCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATT 335

Qy     870 tggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaa 929
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     336 TGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAA 395

Qy     930 aatcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccccctat 983
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     396 AATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATACAACCCCCTAT 449


RESULT  25
HPY269891
LOCUS       HPY269891     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate AU5
            from Australia.
ACCESSION   AJ269891
VERSION     AJ269891.1  GI:5921359
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
```



```
                     /isolate="AU5"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56453.1"
                     /db_xref="GI:5921360"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAPLKVDNAVASYDPDQK
                     PIIDKNDRDNRQAPDGISQLREEYSNKAIKNPTKKNQYFSDFIDKSNDLINKDNLIDV
                     ESSTKSFQKFGDQRYRIPTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      168 a      95 c      69 g     117 t
ORIGIN


  Query Match           3.3%;  Score 194;  DB 2;  Length 449;
  Best Local Similarity   99.6%;  Pred. No. 6.8e-81;
  Matches  244;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     739 tccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaat 798
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 TCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCGAT 264

Qy     799 aagagcaatgatttaatcaacaaagacaatctcattgatgtagaatctttccacaaagagc 858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     265 AAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGC 324

Qy     859 tttcagaaatttggggatcagcgttaccgaatttttcacaagttgggtgtcccatcaaaac 918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325 TTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAAC 384

Qy     919 gatccgtctaaaatcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccc 978
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     385 GATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATACAACCC 444

Qy     979 cctat 983
           |||||
Db     445 CCTAT 449


RESULT  26
AF190659
LOCUS       AF190659     1233 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain HUP B5 cag pathogenicity island
            polymorphic right end, type Ib motif; insertion sequence miniIS605,
            complete aequence; and glutamate racemase (glr) gene, partial cds.
ACCESSION   AF190659
```



```
VERSION     AF190659.1  GI:10567545
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 1233)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 1233)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .1233
                     /organism="Helicobacter pylori"
                     /strain="HUP B5"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="Spain"
                     /note="isolated from a gastric ulcer patient; polymorphic
                     region at right end of the cag pathogenicity island; type
                     Ib motif"
     misc_feature    31. .430
                     /note="similar to Helicobacter pylori insertion sequence
                     IS606 presented in GenBank Accession Number U95957"
     source          66. .333
                     /organism="Helicobacter pylori"
                     /strain="HUP B5"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /insertion_seq="miniIS605"
                     /country="Spain"
                     /note="isolated from a gastric ulcer patient; polymorphic
                     region at right end of the cag pathogenicity island; type
                     Ib motif"
     gene            492. .871
                     /gene="putative DNA helicase"
                     /note="truncated; similar to Helicobacter pylori strain
```



```
                     26695 putative helicase gene HP0548"
     variation       502. .503
                     /gene="putative DNA helicase"
                     /note="46 bp deletion compared to Helicobacter pylori
                     strain NCTC11638"
     repeat_region   1027. .1233
                     /note="defines the terminus of the cag pathogenicity
                     island"
                     /rpt_type=direct
     gene            complement(1027. .>1233)
                     /gene="glr"
                     /note="flanks cag pathogenicity island"
     CDS             complement(1027. .>1233)
                     /gene="glr"
                     /note="similar to Helicobacter pylori strain 26695
                     glutamate racemase HP0549"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG18482.1"
                     /db_xref="GI:10567546"
                     /translation="AQKIEGYFMGHFALPAPPLLIHSGDAIVEYLQKNYTLKKNAHAF
                     PKVEFHASGDVIWLERQAKEWLKL"
BASE COUNT      391 a     237 c     240 g     365 t
ORIGIN


  Query Match           3.1%;  Score 185;  DB 1;  Length 1233;
  Best Local Similarity   99.3%;  Pred. No. 1.2e-76;
  Matches  285;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    4042 ctccttagcagaaatcccaatcgtctttagtatttgggatgaatgctaccaattcatggt 4101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 CTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGATGAATGCTACCAATTCATGGT 127

Qy    4102 atcatatcccctacattcgtatctagcgtaggaagtgtgcaaagttacgccttttggaga 4161
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 ATCATGTCCCCATACATTCGTATCTAGCGTAGGAAGTGTGCAAAGTTACGCCTTTGGAGA 187

Qy    4162 tatgatgtgtgagacctgtagggaatgcgttggagctcaaactctgtaaaatccctatta 4221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     188 TATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCAAACTCTGTAAAATCCCTACTA 247

Qy    4222 tagggacacagagtgagaaccaaactctccctacgggcaacatcagcctaggaagcccaa 4281
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 TAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCAACATCAGCCTAGGAAGCCCAA 307

Qy    4282 tcgtctttagcggttgggcacttcaccttaaaatatcccgacagaca 4328
           |||||||||||||||||||||||||||||||||||||||||||||||
Db     308 TCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCCGACAGACA 354
```

```
RESULT 27
AE000569
LOCUS       AE000569    11141 bp    DNA             BCT       06-APR-1999
DEFINITION  Helicobacter pylori 26695 section 47 of 134 of the complete genome.
ACCESSION   AE000569 AE000511
VERSION     AE000569.1  GI:2313663
KEYWORDS    .
SOURCE      Helicobacter pylori 26695.
  ORGANISM  Helicobacter pylori 26695
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 11141)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and
            Venter,J.C.
  TITLE     The complete genome sequence of the gastric pathogen Helicobacter
            pylori
  JOURNAL   Nature 388 (6642), 539-547 (1997)
  MEDLINE   97394467
  REMARK    Erratum:[[published erratum appears in Nature 1997 Sep
            25:389(6649):412]]
REFERENCE   2  (bases 1 to 11141)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C.,
            Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M.,
            Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 11141)
  AUTHORS   White,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11141
                     /organism="Helicobacter pylori 26695"
                     /strain="26695"
                     /db_xref="taxon:85962"
     gene            447. .4007
```



```
                     /gene="HP0547"
     CDS             447. .4007
                     /gene="HP0547"
                     /note="similar to GP:1752700 percent identity: 92.87;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="cag pathogenicity island protein (cag26)"
                     /protein_id="AAD07614.1"
                     /db_xref="GI:2313664"
                     /translation="MTNETIDQTRTPDQTQSQTAFDPQQFINNLQVAFIKVDNVVASF
                     DPDQKPIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFIDKSNDLINKD
                     NLIDVESSTKSPQKFGDQRYQIFTSWVSHQKDPSKINTRSIRNFMENIIQPPIPDDKE
                     KAFFLKSAKQSFAGIIIGNQIRTDQKFWGVFDESLKERQEAENGGPTGGDWLLTFLS
                     FIFNKKQSSDVKEAINQEPVPHVQPDIATTTTDIQGLPPEARDLLDERGNPSKYIDH
                     MEMLDVEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLYAGNGGT WFH
                     DWNATVGYKDQQGNNVATLINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQLTGSQRA
                     LSQEEIRNKVDFMEFLAQNNTKLDNLSEKEKEKFQNEIEDFQKDSKAYLDALGNDRIA
                     FVSKKDTKHSALITEFNNGDLSYTLKDYGKKADKALDREKNVTLQGSLKHDGVMFVDY
                     SNFKYTNASKNPNKGVGATNGVSHLEAGPNKVAVFNLPDLNNLAITSFVRRNLENKLT
                     AKGLSLQEANKLIKDPLSSNKELAGKALNFNKAVAEAKSTGNYDEVKKAQKDLEKSLR
                     KREHLEKEVEKKLESKSGNKNKMEAKAQANSQKDEIFALINKEANRDARAIAYTQNLK
                     GIKRELSDKLEKISKDLKDFSKSPDEFKNGKNKDFSRAEETLKALKGSVKDLGINPEW
                     ISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKVTDKVDNLNQAVSV
                     AKAMGDFSRVEQVLADLKNFSKEQLAQQAQKNEDFNTGKNSELYQSVKNSVNKTLVGN
                     GLSGIEATALAKNFSDIKKELNEKFKNFNNNNNGLKNSTEPIYAKVNKKKTGQVASPE
                     EPIYTQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGLSASPEPIY
                     ATIDDLGGPFPLKRHDKVDDLSKVGRSRNQELAQKIDNLNQAVSEAKAGFFGNLEQTI
                     DKLKDSTKKNVMNLYVESAKKVPASLSAKLDNYAINSHTRINSNIQNGAINEKATGML
                     TQKNPEWLKLVNDKIVAHNVGSVSLSEYDKIGFNQKNMKDYSDSFKFSTKLNNAVKDI
                     KSGFTHFLANAFSTGYYCLARENAEHGIKNVNTKGGFQKS"
     gene            4136. .4963
                     /gene="HP0548"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     GB:L77117 SP:Q57568 PID:1590880 percent identity: 38.82;
                     identified by sequence similarity; putative; DNA helicase,
                     putative"
     gene            complement(5109. .5876)
                     /gene="HP0549"
     CDS             complement(5109. .5876)
                     /gene="HP0549"
                     /note="similar to GB:U12405 PID:520574 SP:P52974 percent
                     identity: 38.61; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate racemase (glr)"
                     /protein_id="AAD07615.1"
                     /db_xref="GI:2313665"
                     /translation="MKIGVFDSGVGGFSVLKSLLKAQLFDEIIYYGDSARVPYGTKDP
```



```
                     TTIKQFGLEALDFFKPHQIKLLIVACNTASALALEEMQRHSKIPVVGVIEPSILAIKR
                     QVKDKNAPILVLGTKATIQSNAYDNALKQQGYLNVSHLATSLFVPLIEESILEGELLE
                     TCMRYYPTPLEILPEVVILGCTHFPLIAQKIEGYFMEHFALSTPPLLIHSGDAIVEYL
                     QQNYALKKNACAFPKVEFHASGDVVWLEKQAKEWLKL"
     gene            complement(5914. .7230)
                     /gene="HP0550"
     CDS             complement(5914. .7230)
                     /gene="HP0550"
                     /note="similar to GB:L42023 SP:P44619 PID:1003483
                     PID:1222221 PID:1204548 percent identity: 56.59;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcription termination factor Rho (rho)"
                     /protein_id="AAD07616.1"
                     /db_xref="GI:2313666"
                     /translation="MNENAPTHKSSHKVRTHTPVSGYHIEDLRTYPTEKLLEIANKLK
                     VENPQEFKRQDLMFEILKTQVTQGGYILFTGILEIMPDGYGFLRGFDGSFSDGHNDTY
                     VSPSQIRRFALRNGDIVTGQVRSPKDQEKYYALLKIEAINYSPSDEIRNRPLFDNLTP
                     LPPDEQIKLEYEPTKVTGRMLDLFSPVGKGQRALIVAPPRTGKTELMKELAQGITSNH
                     PEVELIILLVDERPEEVTDMQRSVKGQVFSSTFDLPANNHIRIAELVLERAKRRVEMG
                     KDVVLLDSITRLARAYNAVTPSSGKVLSGGVDANALHRPKRFFGAARNIEEGGSLTI
                     IATALIETGSRMDEVIFEEFKGTGNSEIVLARNIADRRIYPAFDILKSGTRKDNILLG
                     KDRLTKVWVLRNVMQQMDDIEALSFVYSKMQQTKDNEEFLNLMNEK"
     gene            7494. .7697
                     /gene="HP0551"
     CDS             7494. .7697
                     /gene="HP0551"
                     /note="similar to GB:M97678 SP:Q03223 PID:143435
                     PID:853770 GB:AL009126 percent identity: 49.25; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosomal protein L31 (rpl31)"
                     /protein_id="AAD07617.1"
                     /db_xref="GI:2313667"
                     /translation="MKKGIHPEYIPCKVTCVTSGKEIEVLSTKPEMRIDISSFCHPFY
                     TGSDKIADTAGRVEKFKQRYNLK"
     gene            7722. .8585
                     /gene="HP0552"
     CDS             7722. .8585
                     /gene="HP0552"
                     /note="similar to SP:P45528 PID:606086 GB:U00096
                     PID:1789535 percent identity: 37.78; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD07618.1"
                     /db_xref="GI:2313668"
                     /translation="MLYFLPTPIGNLADITLRALEVLERCEVFLCEDTRVSKRLLHLL
```



```
                     AQNPIISHSFPNIATKKREFIAPHSHNDQEFLNQIKPSFFDKEIAVMSDAGMPSLSDP
                     GMSLAAYALKHNIQYDVLPGANALTTAFCASGFLEGRFFYAGFLPHKSKERRLKIAKI
                     LNALAYLEKTPVVFYESPHRLLETLKDLNDLAKGMHLFAAKELTKLHQQYYLGEVSQ
                     IIERLQQSTIQGEWVLVLLNEKKIEPCHGLSALLELDLPPRIKAKIEAVMTQKNAKEL
                     YFQRLLEEKNQ"
     gene            8598. .9281
                     /gene="HP0553"
     CDS             8598. .9281
                     /gene="HP0553"
                     /note="similar to GB:L42023 SP:P44906 PID:1005923
                     PID:1220960 PID:1205109 percent identity: 30.04;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD07619.1"
                     /db_xref="GI:2313669"
                     /translation="MQAVIYGKQVIMHLLNSHQEKLQEIYLSKEIDKKLFFALKKACP
                     NIIKVDNKKAQSLAKGGNHQGVLAKVELPLAVSLKEVKKAQKLLVLCGITDVGNIGGI
                     FRSAYCLGMGGVILDFAKELAYEGIVRSSLGLWYDLPFSVMPNTLDLINELKTSGFLC
                     LGASMQGSSQIENLSLKKCALPLGSEHEGLSKKILAKMDTILSVKMRRDFDSLNVSVA
                     AGILMDKIN"
     gene            9294. .10259
                     /gene="HP0554"
     CDS             9294. .10259
                     /gene="HP0554"
                     /note="hypothetical protein: identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0554"
                     /protein_id="AAD07620.1"
                     /db_xref="GI:2313670"
                     /translation="MEQNKKSLENLDLSDVQNISKDISGATLEELSLKNLDKNLQILK
                     EVGVAEICKATKIASKNIHSILEKRYESLSRVHARGFIQILEREYKIDLSAWMKEFDK
                     ACTFKEGVSEEQNQETDPEEKTKNPLKVEIDYSINQANIKLSKGLSKWKPPVLVLGVV
                     VIVLAVVIIQNSSSLKEERGQESAIKSGTKKSFFNKANPIEENKPEPTPKLEEKPKEQ
                     DKQEKEAIKEDPNTIYIIPKKDIWVEVIDLDEKKNSFQKVFKKNYSLETKNHRLLLRF
                     GHGHLSLKNNHQEQDYNDSKTRRFLYEPNKGLTLINEAQYKELQR"
     gene            10256. .11077
                     /gene="HP0555"
     CDS             10256. .11077
                     /gene="HP0555"
                     /note="hypothetical protein: identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0555"
                     /protein_id="AAD07621.1"
                     /db_xref="GI:2313671"
```

```
  Query Match             3.1%;  Score 182;  DB 1;  Length 11141;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-75;
  Matches  182;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2707 tcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaaccttaatgca 2766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2634 TCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAAAGTTGAAAACCTTAATGCA 2693

Qy   2767 gctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaagcaaaa 2826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2694 GCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGTAACGCAAGCAAAA 2753

Qy   2827 agcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggataaagtt 2886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2754 AGCGACCTTGAAAATTCCGTTAAAGATGTGATCATCAATCAAAAGGTAACGGATAAAGTT 2813

Qy   2887 ga 2888
          ||
Db   2814 GA 2815


RESULT  28
AF190663
LOCUS       AF190663      787 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain India75A cag pathogenicity island
            polymorphic right end, type IIIb motif; insertion sequence
            miniIS605, complete sequence; and glutamate racemase (glr) gene,
            partial cds.
ACCESSION   AF190663
VERSION     AF190663.1  GI:10567553
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 787)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (29-SEP-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .787
                     /organism="Helicobacter pylori"
                     /strain="India75A"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="India: Calcutta"
                     /note="isolated from a gastric ulcer patient; polymorphic
                     region at right end of the cag pathogenicity island; type
                     IIIb motif; similar to Helicobacter pylori strain Sweden53
                     type IIIa motif presented in GenBank Accession Number
                     AF190662"
     misc_feature    31. .430
                     /note="similar to Helicobacter pylori insertion sequence
                     IS606 presented in GenBank Accession Number U95957"
     source          66. .333
                     /organism="Helicobacter pylori"
                     /strain="India75A"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /insertion_seq="miniIS605"
                     /country="India: Calcutta"
                     /note="isolated from a gastric ulcer patient; polymorphic
                     region at right end of the cag pathogenicity island; type
                     IIIb motif; similar to type IIIa motif"
     misc_feature    495. .744
                     /note="this sequence differs from type I strains such as
                     J99 and Gambia which contain a putative helicase gene at
                     this location"
     gene            complement(753. .>787)
                     /gene="glr"
                     /note="flanks cag pathogenicity island"
     CDS             complement(753. .>787)
                     /gene="glr"
                     /note="similar to Helicobacter pylori strain 26695
                     glutamate racemase HP0549"
                     /codon_start=3
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG18486.1"
                     /db_xref="GI:10567554"
                     /translation="EKQAKEWLKL"
BASE COUNT      232 a    137 c    173 g    245 t
ORIGIN


  Query Match             3.1%;  Score 181;  DB 1;  Length 787;
  Best Local Similarity   99.6%;  Pred. No. 1e-74;
```



```
  Matches  231;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    167 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 226

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    227 AACTCTGTAAAATCCCTACTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 286

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    287 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 346

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcct 4372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCT 398


RESULT  29
HPY269879
LOCUS       HPY269879     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate PO9
            from Portugal.
ACCESSION   AJ269879
VERSION     AJ269879.1  GI:5921335
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="PO9"
                     /db_xref="taxon:210"
                     /country="Portugal"
     gene            1. .449
                     /gene="cagA"
```



```
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56447.1"
                     /db_xref="GI:5921336"
                     /translation="MTNETIDQQPQTEAAFNPQQFINNLQVAFLKLDNAVASFDPDQK
                     PIVDKNDRDNRQAPDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDALIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      162 a     96 c     72 g    119 t
ORIGIN


  Query Match             3.0%;  Score 179;  DB 2;  Length 449;
  Best Local Similarity   99.1%;  Pred. No. 9.3e-74;
  Matches  329;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    652 gatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaa 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 GATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAT 177

Qy    712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 GGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAG 237

Qy    772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    238 AATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACGCTCTC 297

Qy    832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 ATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATT 357

Qy    892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 TTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGA 417

Qy    952 aattttatggaaaatatcatacaacccccctat 983
          |||||||||||||||||||||||||||||||||
Db    418 AATTTTATGGAAAATATCATACAACCCCCTAT 449


RESULT  30
HPY269882
LOCUS       HPY269882     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            NL82 from The Netherlands.
ACCESSION   AJ269882
VERSION     AJ269882.1  GI:5921341
KEYWORDS    cagA gene; CagA protein.
```

```
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="NL82"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56496.1"
                     /db_xref="GI:5921342"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFLKLDNAVASFDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFIDKSNDLINKDNLIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      164 a      96 c      71 g     118 t
ORIGIN
```

```
  Query Match             3.0%;  Score 179;  DB 2;  Length 449;
  Best Local Similarity   99.1%;  Pred. No. 9.3e-74;
  Matches  329;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    652 gatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaa 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 GATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAT 177

Qy    712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 GGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAG 237

Qy    772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
          |
```



```
          ||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db    238 AATCAGTATTTTTCAGACTTTATCGATAAGAGCAACGATTTAATCAACAAAGACAATCTC 297

Qy    832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 ATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATT 357

Qy    892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 TTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGA 417

Qy    952 aattttatggaaaatatcatacaacccccctat 983
          |||||||||||||||||||||||||||||||||
Db    418 AATTTTATGGAAAATATCATACAACCCCCTAT 449
```

```
RESULT  31
HPY239691
LOCUS       HPY239691     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain 5596.
ACCESSION   AJ239691
VERSION     AJ239691.1  GI:4467700
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="5596"
                     /db_xref="taxon:210"
                     /country="Gambia"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
```



```
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37802.1"
                     /db_xref="GI:4467701"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVASYDPDQKPIVDKNDRDNRQ
                     AFEGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTKSFQKFGD
                     QRYRIFTSWVSHQNDPSKINTRSIQNFMEHTIQPP"
BASE COUNT      148 a      86 c      69 g     110 t
ORIGIN
```

```
  Query Match             3.0%;  Score 175;  DB 2;  Length 413;
  Best Local Similarity   98.9%;  Pred. No. 7.5e-72;
  Matches  375;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    571 gaagcggcttttaacccgcagcaatttatcaataatcttcaagtagcttttcttaaagtt 630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGTT 60

Qy    631 gataacgctgtcgcttcatacgatcctgatcaaaaaccaatcgttgataagaacgatagg 690
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATAATGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGG 120

Qy    691 gataacaggcaagcttttgaaggaatctcgcaattaagggaagaatactccaataaagcg 750
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    121 GATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCG 180

Qy    751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 240

Qy    811 ttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaattt 870
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| |||
Db    241 TTAATCAACAAAGACAATCTCATTGATGTGGAATCTTCCACAAAGAGCTTTCAGAAGTTT 300

Qy    871 ggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaa 930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAA 360

Qy    931 atcaacacccgatcgatcc 949
          |||||||||||||||||||
Db    361 ATCAACACCCGATCGATCC 379
```

```
RESULT  32
HPY269888
LOCUS       HPY269888     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            NL45 from The Netherlands.
ACCESSION   AJ269888
VERSION     AJ269888.1  GI:5921353
```



```
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="NL45"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56494.1"
                     /db_xref="GI:5921354"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSNFINKSNDLINKDNLIDV
                     ESSKKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      169 a      97 c      69 g     114 t
ORIGIN
```

```
  Query Match             3.0%;  Score 175;  DB 2;  Length 449;
  Best Local Similarity   98.8%;  Pred. No. 7.5e-72;
  Matches  425;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 ACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAG 79

Qy    614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 TGGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCG 139
```

```
Qy    674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    140 TTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGATGGAATCTCGCAATTAAGGGAAG 199

Qy    734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 AATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAAACTTTA 259

Qy    794 tcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaa 853
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||| ||
Db    260 TCAATAAGAGCAATGATCTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCAAAA 319

Qy    854 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 AGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATC 379

Qy    914 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaaatatcatac 973
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 AAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATAC 439

Qy    974 aaccccctat 983
          ||||||||||
Db    440 AACCCCCTAT 449


RESULT  33
HPY269883
LOCUS       HPY269883     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            PE9012 from Peru.
ACCESSION   AJ269883
VERSION     AJ269883.1  GI:5921343
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
```



```
                     /organism="Helicobacter pylori"
                     /isolate="PE9012"
                     /db_xref="taxon:210"
                     /country="Peru"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56450.1"
                     /db_xref="GI:5921344"
                     /translation="MTNETIDQQPQTEAAPNPQQPINNLQVAFLKLDNAVASFDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSQINVQKIRNFMENIIQPP"
BASE COUNT      166 a     94 c     70 g     119 t
ORIGIN


  Query Match             2.9%;  Score 174;  DB 2;  Length 449;
  Best Local Similarity   99.3%;  Pred. No. 2.2e-71;
  Matches  274;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    652 gatcctgatcaaaaaccaatcgttgatagaagaacgatagggataacagggcaagcttttgaa 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 GATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAT 177

Qy    712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 GGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAG 237

Qy    772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 AATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTC 297

Qy    832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 ATTGATGTGGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATT 357

Qy    892 ttcacaagttgggtgtcccatcaaaacgatccgtct 927
          |||||||||||||||||||||||||||||||||||
Db    358 TTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCT 393


RESULT  34
AF202220
LOCUS       AF202220      287 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain India7A CagA (cagA) gene, partial cds.
ACCESSION   AF202220
VERSION     AF202220.1  GI:7229431
```



```
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
            Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
            Nair,G.B. and Berg,D.E.
  TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
            India
  JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
  MEDLINE   20270153
   PUBMED   10809703
REFERENCE   2  (bases 1 to 287)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Chowdhury,A., Datta,S., Nair,G.B.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) Molecular Microbiology, Washington
            University, 660 South Euclid Ave., St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .287
                     /organism="Helicobacter pylori"
                     /strain="India7A"
                     /db_xref="taxon:210"
     gene            <1. .>287
                     /gene="cagA"
     CDS             <1. .>287
                     /gene="cagA"
                     /note="similar to Helicobacter pylori 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="CagA"
                     /protein_id="AAF42822.1"
                     /db_xref="GI:7229432"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT      105 a     52 c     65 g     65 t
ORIGIN


  Query Match             2.9%;  Score 169;  DB 1;  Length 287;
  Best Local Similarity   99.5%;  Pred. No. 5.5e-69;
  Matches  219;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db    128 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 187

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaaatagatttcatggaattt 1740
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    188 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCGAAACAAAATAGATTTCATGGAATTT 247

Qy   1741 cttgcacaaaataatgctaaattagacaacttgagcgaga 1780
          |||||||||||||||||||||||||||||||||||||||
Db    248 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCGAGA 287


RESULT  35
AF289438
LOCUS       AF289438      672 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR8 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289438
VERSION     AF289438.1  GI:9964230
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 672)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Helicobacter pylori"
                     /isolate="CR8"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>672
                     /gene="cagA"
     CDS             <1. .>672
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
```

/product="cytotoxin associated protein A"
/protein_id="AAG09887.1"
/db_xref="GI:9964231"
/translation="MFSDIKKELNAKLFGNPNNNNNNGLKNEPIYAKVNKKKTGQVAS
PEEPIYAQVAKKVNAKIDRFNQIASGLGGVGQAAGFPPLKRHDKVDDLSKVGRSVSPEP
IYATIDDLGGPFPLKRHDKVGDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVGDLS
KVGLSREQQLAQKIDNLNQAVSEAKAGFFGNLEQTIDKLKDSTKHNPMNLWVESAKKV
PASLSA"
BASE COUNT 231 a 121 c 150 g 170 t
ORIGIN

Query Match 2.8%; Score 167; DB 2; Length 672;
Best Local Similarity 100.0%; Pred. No. 4.7e-68;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3422 aattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggttttt 3481
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 506 AATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTTT 565

Qy 3482 ttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatga 3541
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 566 TTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATGA 625

Qy 3542 atctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcg 3588
|||||||||||||||||||||||||||||||||||||||||||||||
Db 626 ATCTATGGGTTGAAAGTGCAAAAAAAGTACCTGCTAGTTTGTCAGCG 672

RESULT 36
AE001483
LOCUS AE001483 10781 bp DNA BCT 20-JAN-1999
DEFINITION Helicobacter pylori, strain J99 section 44 of 132 of the complete genome.
ACCESSION AE001483 AE001439
VERSION AE001483.1 GI:4155034
KEYWORDS .
SOURCE Helicobacter pylori J99.
ORGANISM Helicobacter pylori J99
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 10781)
AUTHORS Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C., Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G., Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C., Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E., Vovis,G.F. and Trust,T.J.
TITLE Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
JOURNAL Nature 397 (6715), 176-180 (1999)
MEDLINE 99120557

REMARK Erratum:[[published erratum appears in Nature 1999 Feb 25;397(6721):719]]
REFERENCE 2 (bases 1 to 10781)
AUTHORS King,B.L., Alm,R.A. and Trust,T.J.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA
COMMENT Address all correspondence to: hpmarcb.us.astra.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith, Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: http://www.astra-boston.com/hpylori).
FEATURES Location/Qualifiers
source 1. .10781
/organism="Helicobacter pylori J99"
/strain="J99"
/db_xref="taxon:85963"
gene 417. .3920
/gene="cagA"
/note="jhp0495"
CDS 417. .3920
/gene="cagA"
/note="similar to H. pylori 26695 gene HP0547"
/codon_start=1
/transl_table=11
/product="cag island protein, CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN"
/protein_id="AAD06073.1"
/db_xref="GI:4155035"
/translation="MTNEAINQQPQTEAAFNPQQFINNLQVAFIKVDNVVASFDPNQK
PIVDKNDRDNRQAFEKISQLREEPANKAIKNPTKKNQYFSSFISKSNDLIDKDNLIDT
GSSIKSFQKFGTQRYQIFMNWVSHQNDPSKINTQKIRGFMENIIQPPISDDKEKAEFL
RSAKQAFAGIIIGNQIRSDQKPMGVFDESLKERQEAEKNGEPNGDPTGGDWLDIFLSF
VFNKKQSSDLKETLNQEPVPHVQPDVATTTTDIGSLPPEARDLLDERGNFSKFTLGDM
NMLDVEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLYGNNGGPEARHD
WNATVGIYKNQRGDNVATLINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQLTGSQRAL
SQEEIQNKVDPMEPLAQNNAKLDNLSKKEKEKFQNEIEDFQKDSKAYLDALGNDHIAF
VSKKDKHLALVAEFGNGELSYTLKDYGKKADKALDREAKTTLQGSLKHDGVMFVDYS
NFKYTNASKSPDKGVGATNGVSHLEAGFSKVAVFNLPNLNNLAITSVVRQDLEDKLIA
KGLSPQEANKLVKDFLSSNKELVGKALNFNKAVAEAKNTGNYDEVKQAQKDLEKSLKK
RERLEKDVAKNLESKSGNKNKMEAKSQANSQKDEIFALINKEANRDARAIAYAQNLKG
IKRELSDKLENINKDLKDFSKSFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWI

SKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSIKDVIINQKITDKVDNLNQAVSVA
KATGDFSGVEQALADLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGVNGTLVGNG
LSKAEATTLSKNFSDIKKELNAKLGNFNNNNNNGLENSTEPIYTQVAKKVKAKIDRLD
QIASGLGDVGQAASFLLKRHDKVDDLSKVGLSANHEPIYATIDDLGGPPPLKRHDKVD
DLSKVGLSREQKLTQKIDNLNQAVSEAKASHFDNLDQMIDKLKDSTKNVVNLYVESA
KKVPTSLSAKLDNYATNSHTRINSNVKNGTINEKATGMLTQKNSEWLKLVNDKIVAHN
VGSAPLSAYDKIGFNQKNMKDYSDSFKFSTRLSNAVKDIKSGFVQPLTNIPSMGSYSL
MKASVEHGVKNTNTKGGFQKS"
gene complement(4739. .5506)
/gene="murI"
/note="jhp0496"
CDS complement(4739. .5506)
/gene="murI"
/note="similar to H. pylori 26695 gene HP0549"
/codon_start=1
/transl_table=11
/product="GLUTAMATE RACEMASE"
/protein_id="AAD06074.1"
/db_xref="GI:4155036"
/translation="MKIGVFDSGVGGFSVLKSLLKARLFDEIIYYGDSARVPYGTKDP
TTIKQFGLEALDFFKPHEIELLIVACNTASALAEEMQKYSKIPIVGVIEPSILAIKR
QVEDKNAPILVLGTKATIQSNAYDNALKQQGYLNISHLATSLFVPLIEESILEGELLE
TCMHYYFTPLEILPEVIILGCTHFPLIAQKIEGYFMGHFALPTPPLLIHSGDAIVEYL
QQKYALKNNACTFPKVEPHASGDVIWLERQAKEWLKL"
gene complement(5546. .6862)
/gene="rho"
/note="jhp0497"
CDS complement(5546. .6862)
/gene="rho"
/note="similar to H. pylori 26695 gene HP0550"
/codon_start=1
/transl_table=11
/product="TRANSCRIPTION TERMINATION FACTOR"
/protein_id="AAD06075.1"
/db_xref="GI:4155037"
/translation="MNENAPTHKSSHKVKTHTPVSGYHIEDLRTYPTEKLLEIANKLK
VENPQEFKRQDLMFEILKTQVTQGGYILFTGILEIMPDGYGFLRGFDGSFSDGHNDTY
VSPSQIRRFALRNGDIVTGQVRSPKDQEKYYALLKIEAINYLPSDEIKNRPLFDNLTP
LFPDEQIKLEYEPTKVTGRMLDLFSPVGKGQRALIVAPPRTGKTELMKELAQGITSNH
PEVELIILLVDERPEEVTDMQRSVKGQVFSSTFDLPANNHIRIAELVLERAKRRVEMG
KDVVLLDSITRLARAYNATPSSGKVLSGGVDANALHRPKRFFGAARNIEEGGSLTI
IATALIETGSRMDEVIFEEFKGTGNSEIVLARNIADRRIYPAFDILKSGTRKDNILLG
KDRLTKVWVLRNVMQQMDDIEALSFVYSKMQQTKDNEEFLNLMNEK"
gene 7127. .7330
/gene="rpmE"
/note="jhp0498"
CDS 7127. .7330
/gene="rpmE"
/note="similar to H. pylori 26695 gene HP0551"
/codon_start=1
/transl_table=11

/product="50S RIBOSOMAL PROTEIN L31"
/protein_id="AAD06076.1"
/db_xref="GI:4155038"
/translation="MRKKGIHPEYIPCKVTCVTSGKEIEVLSTKPEMRIDISSFCHPFY TGSDKIADTAGRVEKFKQRYNLK"
gene 7354. .8223
/gene="jhp0499"
CDS 7354. .8223
/gene="jhp0499"
/note="similar to H. pylori 26695 gene HP0552"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06077.1"
/db_xref="GI:4155039"
/translation="MLYFLPTPIGNLADITLRTLEVLERCEVFLCEDTRVSKRLLHLL AKNPIISHSFPNIAAKKREFIAFHSHNDQEFLNQIEPSFFDKEIAVMSDAGMPSLSDP GMSLVAYALKHNLQYDVLPGANALTTAFCASGFLEGRFFYAGFLPHKSKERRLRTIKI LNALAYLEEKTPVVFYESPHRLLETLRDLNDLAQGMHLFAAKELTKLHQQYYLGEISQ IMTQLQKSNIQGEWVLVLLNEKKIEPSMGLSALLELDLPPKIKAKMEAAMTQKNAKEL YFQRLLEEKKQCD"
gene 8237. .8920
/gene="jhp0500"
CDS 8237. .8920
/gene="jhp0500"
/note="similar to H. pylori 26695 gene HP0553"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06078.1"
/db_xref="GI:4155040"
/translation="MQAVIVGKQVITHLLNSHHEKLQEIYLSKEIDKKLFFALKKACP NIIKVDNKKAQSLAKGGNHQGVLAKVELPLAASLKEIKKAQKLLVLCGITDVRNIGGI FRSAYCLGMDGVILDFARAEFAYEGIVRSSLGLMYDLPFSVMPNTLDLINELKTSGFLC LGASMQGSSQAENLSLKKCALPLGSEHEGLSKKILAKMDAILSVKMRRDFDSLNVSVA AGILMDKIN"
gene 8933. .9898
/gene="jhp0501"
CDS 8933. .9898
/gene="jhp0501"
/note="similar to H. pylori 26695 gene HP0554"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06079.1"
/db_xref="GI:4155041"
/translation="MEQNKKSLENLDLSDVQNISKDISGTALEELSLKNLDKNLQILK EVGAAEICKATKIASKNIHSILEKRYESLSRVHARGFIQILEHEYKIDLSAWVKEFDK VCVFKEGVGEEQKQETSPEETAKKPLKVELDYSINQANTSLSKKSSKWKPFVIVLGVV VIILVVVIIQNSSSLKEERKQERAIKPDTKNNSPNETNPTEEKKLEPTPKLEEKHKEQ DKQGKEAIKENPNTIYIIPKRDIWVEVIDLDEKKNSFQKVFKKSYPLEAKNHRLLLRF

```
                     GHGHLILKNNHQEQDYNDSKTRRFLYEPNKGLTLINEAQYKALQQ"
     gene            9895. .10716
                     /gene="jhp0502"
     CDS             9895. .10716
                     /gene="jhp0502"
                     /note="similar to H. pylori 26695 gene HP0555"
                     /codon_start=1
                     /transl_table=11
                     /product="putative"
                     /protein_id="AAD06080.1"
                     /db_xref="GI:4155042"
                     /translation="MNKLFLAFIVGGMLLSADALNDKIENLMGERSYHMNKLFLERLF
                     KNRKDFYEMGRLDSLKLLNTLKENGLLSFNFDKPSVLKITFKASSNPLAFAKSINNSL
                     NMMGYSYVLPIRMQSSSGENVFSYELKTEYVLDPNILIETMKRHGFDFMDIRRVSLKE
                     WEYDFALQKIKLPNARALVLSSDPVEFKEASGKYWLSVNQNAYLKISSNNPLWQPKII
                     FYDENLKIIQIIAKENRQQEIALNLLDGVRFIHITDAKNPIILKNGISVVFDAMP"
BASE COUNT     3585 a    1918 c    2165 g    3113 t
ORIGIN
```

```
  Query Match             2.8%;  Score 166;  DB 1;  Length 10781;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-67;
  Matches  216;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    2628 taaatcttttgatgaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaac 2687
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2522 TAAATCTTTTGATGAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGCAGAAGAAAC 2581

Qy    2688 actaaaagcccttaaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaa 2747
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    2582 ACTAAAAGCCCTTAAAGGCTCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAA 2641

Qy    2748 agttgaaaaccttaatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcag 2807
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2642 AGTTGAAAACCTTAATGCAGCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTCAG 2701

Qy    2808 caagggtaacgcaagcaaaaagcgaccttgaaaattcc 2844
           ||||||||||||||||||||||||||||||||||||||
Db    2702 CAAGGTAACGCAAGCAAAAAGCGACCTTGAAAATTCC 2738
```

```
RESULT  37
AF289406
LOCUS       AF289406      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR9 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289406
VERSION     AF289406.1  GI:9964166
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
```



```
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES              Location/Qualifiers
     source           1. .291
                      /organism="Helicobacter pylori"
                      /isolate="CR9"
                      /db_xref="taxon:210"
                      /country="Costa Rica"
                      /note="isolated from a patient with gastric
                      adenocarcinoma"
     gene             <1. .>291
                      /gene="cagA"
     CDS              <1. .>291
                      /gene="cagA"
                      /note="CagA"
                      /codon_start=1
                      /transl_table=11
                      /product="cytotoxin associated protein A"
                      /protein_id="AAG09855.1"
                      /db_xref="GI:9964167"
                      /translation="IKNPTKKNQYFSDFINKSNDLINKDALIDVESSTKSPQKFGDQR
                      YRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature     115. .135
                      /gene="cagA"
                      /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      108 a      63 c      44 g      76 t
ORIGIN
```

```
  Query Match             2.7%;  Score 158;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-64;
  Matches  158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     828 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 TCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCG 137

Qy     888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
```



```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    138 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 197

Qy     948 ccgaaatttttatggaaaaatatcatacaaccccctatcc 985
           |||||||||||||||||||||||||||||||||||||||||
Db     198 CCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 235
```

```
RESULT  38
AF289417
LOCUS       AF289417      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR42 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289417
VERSION     AF289417.1  GI:9964188
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES              Location/Qualifiers
     source           1. .291
                      /organism="Helicobacter pylori"
                      /isolate="CR42"
                      /db_xref="taxon:210"
                      /country="Costa Rica"
                      /note="isolated from a patient with chronic
                      gastritis-associated dyspepsia"
     gene             <1. .>291
                      /gene="cagA"
     CDS              <1. .>291
                      /gene="cagA"
                      /note="CagA"
                      /codon_start=1
                      /transl_table=11
                      /product="cytotoxin associated protein A"
                      /protein_id="AAG09866.1"
                      /db_xref="GI:9964189"
```



```
                      /translation="IKNPTKKNQYFSDFINKSNDLINKDALIDVESSTKSPQKFGDQR
                      YRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature     115. .135
                      /gene="cagA"
                      /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      108 a      63 c      44 g      76 t
ORIGIN
```

```
  Query Match             2.7%;  Score 158;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-64;
  Matches  158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     828 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 TCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCG 137

Qy     888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 197

Qy     948 ccgaaatttttatggaaaaatatcatacaaccccctatcc 985
           |||||||||||||||||||||||||||||||||||||||||
Db     198 CCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 235
```

```
RESULT  39
AF083352
LOCUS       AF083352     4093 bp    DNA             BCT       26-AUG-1998
DEFINITION  Helicobacter pylori cytotoxin associated protein III (cagA) gene,
            complete cds.
ACCESSION   AF083352
VERSION     AF083352.1  GI:3452457
KEYWORDS    .
SOURCE      Helicobacter pylori
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 4093)
  AUTHORS   Graessle,S., Redl,B. and Gapp,G.
  TITLE     Molecular characterization of the cytotoxin associated gene III
            (cagAIII) of Helicobacter pylori
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4093)
  AUTHORS   Graessle,S., Redl,B. and Gapp,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1998) of Microbiology, University of Innsbruck,
            Fritz Pregl Str. 3, Innsbruck A-6020, Austria
FEATURES              Location/Qualifiers
     source           1. .4093
                      /organism="Helicobacter pylori"
```

```
                     /strain="15818"
                     /db_xref="taxon:210"
     gene            181. .3732
                     /gene="cagA"
     CDS             181. .3732
                     /gene="cagA"
                     /note="CagAIII"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein III"
                     /protein_id="AAC32880.1"
                     /db_xref="GI:3452458"
                     /translation="MANETINQQPQTGAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFLDFINKSNGLINKDALIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFL
                     KSAKQSFAGIIIGNQIRTDQKFMGVFDESLKERQEAEKNGGPTGGDWLDIFLSFIFDK
                     KQSSDVKEAINQEPIPHVQPDIATTTTDIQGLPPESRDLLDERGNFSKFTLGDMEMLD
                     VEGVADIDPNYKFNQLLLIHNNALSSVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNA
                     TVGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQLTGSQRALSQE
                     EIQNKIDPMEFLAQNNAKLDNLSVKEKEKFRNEIKDFQKDSKPYLDALGNDRIAFVSK
                     KDTKHSALITEFNKGDLSYTLKDYGKKADRALDREKNVTLQGSLKHDGVMFVDYSNFK
                     YTNASKSPDKGVGATNGVSHLEANLSKVAVFNLPNLNNLAITSYIRRDLEDKLWAKGL
                     SPQETNKLIKDFLNSNKELVEKVSNLNKAVAEAKNTGNYDEVKAQKDLEKSLRKREH
                     LEKEVVKKLENRNDNKNRMEAKAQANSQKDKIFAIINEEASKEARAAAYVQNLKGIRH
                     ELSDKLENINKDLKDFDKSPDEPKNGKNKDPSKTEETLKALKDSVKDLGINPEWISKV
                     ENLNTALNEFKNGKNKDPSKVTQAKSDLENSIKDAIINQKITDKVDNLNQAVSIAKET
                     GNFSGVEQALADLKNPSKEQLAQQAQKNEDFNTGKNSELYQSVKNSVNKTLVGNGLSG
                     IEATALAKKFSDIKKELNKEFKNFNNNNNNGLKNEPIYVKVNKKKTGQVASPEEPIYT
                     QVAKKVNAKIDRLNQIASGLGAVGQAMGFPLKRHDKVDDLSKVGLSANHEPIYATIDD
                     LFGGPFLSFERHDKVDDLSKIGLSRNQELAQKIDNLNQAVSEAKTCHFDNLDQMIDKL
                     KDSTKKNVMNLYVESAKKVPTSLSAKLDNYAINSHTRINSNIQNGAINEKATGMLTQK
                     NPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMKDYSDSFKFSTKLNNAVKDVKSG
                     FTHFLANAFSTGYYCLTEENAEHGIKNVNTKGGFQKS"
BASE COUNT     1502 a    702 c    806 g   1083 t
ORIGIN


  Query Match             2.7%;  Score 158;  DB 1;  Length 4093;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-64;
  Matches  158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    828 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 887
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    474 TCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCG 533

Qy    888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    534 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 593

Qy    948 ccgaaatttttatggaaaatatcatacaaccccctatcc 985
          |||||||||||||||||||||||||||||||||||||||
```



```
Db    594 CCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 631


RESULT  40
AF289422
LOCUS       AF289422      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR51 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289422
VERSION     AF289422.1  GI:9964198
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR51"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09871.1"
                     /db_xref="GI:9964199"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDALIDVESSTKSFQKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPISDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
```



```
BASE COUNT      108 a     60 c     44 g     79 t
ORIGIN


  Query Match             2.6%;  Score 157;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-63;
  Matches  157;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    828 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 887
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     78 TCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCG 137

Qy    888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    138 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 197

Qy    948 ccgaaattttatggaaaatatcatacaaccccctatc 984
          |||||||||||||||||||||||||||||||||||||
Db    198 CCGAAATTTTATGGAAAATATCATACAACCCCCTATC 234


RESULT  41
HPY239684
LOCUS       HPY239684     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain CC28.
ACCESSION   AJ239684
VERSION     AJ239684.1  GI:4467686
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="CC28"
                     /db_xref="taxon:210"
                     /country="South Africa"
```



```
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37795.1"
                     /db_xref="GI:4467687"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVASFDPDQKPIVDKNDRDNRQ
                     AFDGISQLREEYSNKAIKNPTKKNQYFSDFIDKSNDLINKDNLIDVESSTKSFQKFGD
                     QRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      147 a     85 c     70 g    111 t
ORIGIN


  Query Match             2.6%;  Score 157;  DB 2;  Length 413;
  Best Local Similarity   99.2%;  Pred. No. 2.8e-63;
  Matches  257;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    725 taagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtattttt 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    155 TAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTT 214

Qy    785 cagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaat 844
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    215 CAGACTTTATCGATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTAGAAT 274

Qy    845 cttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagttggg 904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    275 CTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGG 334

Qy    905 tgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaa 964
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    335 TGTCACATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAAA 394

Qy    965 atatcatacaaccccctat 983
          . ||||||||||||||||||
Db    395 ATATCATACAACCCCCTAT 413


RESULT  42
HPY269889
LOCUS       HPY269889     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            NL79 from The Netherlands.
ACCESSION   AJ269889
VERSION     AJ269889.1  GI:5921355
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
```

```
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="NL79"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56493.1"
                     /db_xref="GI:5921356"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDRISQLREEYSNKAIKNPTKKNQYFSDFIDKSNDLINKDALIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      166 a    100 c     70 g    113 t
ORIGIN


  Query Match             2.6%;  Score 156;  DB 2;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-63;
  Matches  156;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     828 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     294 TCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCG 353

Qy     888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     354 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 413

Qy     948 ccgaaattttatggaaaatatcatacaaccccctat 983
           ||||||||||||||||||||||||||||||||||||
Db     414 CCGAAATTTTATGGAAAATATCATACAACCCCCTAT 449
```



```
RESULT  43
HPY252970
LOCUS       HPY252970     244 bp    DNA             BCT       08-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
            protein A.
ACCESSION   AJ252970
VERSION     AJ252970.1  GI:5851971
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
REFERENCE   2  (bases 1 to 244)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Helicobacter pylori"
                     /strain="Dutch 79"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .244
                     /gene="cagA"
     CDS             <1. .>244
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB55446.1"
                     /db_xref="GI:5851972"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEILNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       95 a     42 c     50 g     57 t
ORIGIN


  Query Match             2.6%;  Score 154;  DB 2;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-62;
  Matches  154;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy    1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy    1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy    1681 ggctcacaacgagcattaagtcaagaagagatcc 1714
           ||||||||||||||||||||||||||||||||||
Db     145 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCC 178


RESULT  44
AF198472
LOCUS       AF198472      286 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain Guatemala88 cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198472
VERSION     AF198472.1  GI:7248670
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 286)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 286)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
```



```
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .286
                     /organism="Helicobacter pylori"
                     /strain="Guatemala88"
                     /db_xref="taxon:210"
                     /country="Guatemala"
                     /note="European type"
     gene            <1. .>286
                     /gene="cagA"
     CDS             <1. .>286
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43612.1"
                     /db_xref="GI:7248671"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT      104 a     54 c     65 g     63 t
ORIGIN


  Query Match             2.6%;  Score 154;  DB 1;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-62;
  Matches  154;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy    1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 187

Qy    1681 ggctcacaacgagcattaagtcaagaagagatcc 1714
           ||||||||||||||||||||||||||||||||||
Db     188 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCC 221


RESULT  45
AF198479
LOCUS       AF198479      220 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain PeruGJ2 cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198479
VERSION     AF198479.1  GI:7248684
KEYWORDS    .
SOURCE      Helicobacter pylori.
```

```
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 220)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 220)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .220
                     /organism="Helicobacter pylori"
                     /strain="PeruGGJ2"
                     /db_xref="taxon:210"
                     /note="European type; isolated from gastric juice sample
                     from native Peruvian"
     gene            <1. .>220
                     /gene="cagA"
     CDS             <1. .>220
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43619.1"
                     /db_xref="GI:7248685"
                     /translation="KDQQGNNVATIINVHMRNGSGLVIAGGEKGINNPSFYLYKEDQL
                     TGSQRALSQEEIQNKIDFMEFLAQNNAKL"
BASE COUNT       90 a      37 c      41 g      52 t
ORIGIN
```



```
  Query Match             2.6%;  Score 153;  DB 1;  Length 220;
  Best Local Similarity   99.5%;  Pred. No. 2.3e-61;
  Matches  203;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1561 aacaatgtggctacaatattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17 AACAATGTGGCTACAATTAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 76

Qy    1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 136

Qy    1681 ggctccacaacgagcattaagtcaagaagagatccaaaaacaaaatagatttcatggaattt 1740
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 GGTTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 196

Qy    1741 cttgcacaaaaataatgctaaatta 1764
           |||||||||||||||||||||||||
Db     197 CTTGCACAAAATAATGCTAAATTA 220


RESULT  46
AF222809
LOCUS       AF222809     1323 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain India99 CagA (cagA) gene, partial cds.
ACCESSION   AF222809
VERSION     AF222809.1  GI:7021430
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 1323)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
            Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
            Nair,G.B. and Berg,D.E.
  TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
            India
  JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
  MEDLINE   20270153
   PUBMED   10809703
REFERENCE   2  (bases 1 to 1323)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Datta,S., Chowdhury,A., Nair,G.B.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Molecular Microbiology, Washington
            University Medical School, 660 South Euclid Avenue, Saint Louis, MO
            63108, USA
FEATURES             Location/Qualifiers
```



```
     source          1. .1323
                     /organism="Helicobacter pylori"
                     /strain="India99"
                     /db_xref="taxon:210"
     gene            <1. .>1323
                     /gene="cagA"
     CDS             <1. .>1323
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA"
                     /protein_id="AAF35342.1"
                     /db_xref="GI:7021431"
                     /translation="IDLENSVKDVIINQKITDKVDNLSSAVSVAKAMGDFSKVEQALA
                     DLKNFSREQLAQQAQKNESFNTGKKSEIYQSVKNGVNGTLVGNGLSKAEATTLSKNFS
                     DIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKACQAASLEEPIYAQVAKKVNAKID
                     RLNQIASGLGVVGQAAGFPLKRHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKKHD
                     KVDDLSKVGLSASPEPIYATIDDLGGPFPLKKHDKVDDLSKVGLSRNQELAQKIDNLN
                     QAVSEAKAGFFGNLEQTIDKLKDSIKHNPMNLWVESTKKVPASLSAKLDNYATNSHTR
                     INSNIKNGAINEKATGMLTQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMKD
                     YSDSFKFSTKLNNAAKDIKSGFTQFLTNAFSTGYYCLAGENAEHGIKNV"
BASE COUNT      475 a     249 c     276 g     323 t
ORIGIN


  Query Match             2.6%;  Score 152;  DB 1;  Length 1323;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-61;
  Matches  152;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3372 gcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaagaattggctca 3431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     750 GCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAATTGGCTCA 809

Qy    3432 gaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggttttttggcaatct 3491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     810 GAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTTTTTGGCAATCT 869

Qy    3492 agagcaaacgatagacaagctcaaagattcta 3523
           ||||||||||||||||||||||||||||||||
Db     870 AGAGCAAACGATAGACAAGCTCAAAGATTCTA 901


RESULT  47
HPY239692
LOCUS       HPY239692     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain #12.
ACCESSION   AJ239692
VERSION     AJ239692.1  GI:4467702
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
```



```
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="#12"
                     /db_xref="taxon:210"
                     /country="Guatemala"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37803.1"
                     /db_xref="GI:4467703"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVASYDPDQKPIIDKNDRDNRQ
                     AFEGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGD
                     QRYRIFTSWVSHQNDPSKINTRSIRNFMEHIIQPP"
BASE COUNT      147 a      86 c      70 g     110 t
ORIGIN


  Query Match             2.5%;  Score 149;  DB 2;  Length 413;
  Best Local Similarity   99.2%;  Pred. No. 1.8e-59;
  Matches  249;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     586 ccgcagcaatttatcaataatcttcaagtagcttttcttaaagttgataacgctgtcgct 645
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16 CCGCAGCAATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCT 75

Qy     646 tcatacgatcctgatcaaaaaccaatcgttgataagaacgataggggataacaggcaagct 705
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      76 TCATACGATCCTGATCAAAAACCAATCATTGATAAGAACGATAGGGATAACAGGCAAGCT 135

Qy     706 tttgaaggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctacc 765
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    136 TTTGAGGGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACC 195

Qy    766 aaaaagaatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagac 825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 AAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGAC 255

Qy    826 aatctcattga 836
          ||||||||||||
Db    256 AATCTCATTGA 266
```

RESULT 48
AF289420
LOCUS AF289420 291 bp DNA BCT 02-SEP-2000
DEFINITION Helicobacter pylori isolate CR49 cytotoxin associated protein A (cagA) gene, partial cds.
ACCESSION AF289420
VERSION AF289420.1 GI:9964194
KEYWORDS .
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 291)
AUTHORS Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N., Covacci,A. and Megraud,F.
TITLE Composition and gene expression of the cag pathogenicity island in Helicobacter pylori strains isolated from gastric carcinoma and gastritis patients in Costa Rica
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 291)
AUTHORS Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N., Covacci,A. and Megraud,F.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2000) Laboratory of Bacteriology, University of Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES Location/Qualifiers
source 1. .291
/organism="Helicobacter pylori"
/isolate="CR49"
/db_xref="taxon:210"
/country="Costa Rica"
/note="isolated from a patient with chronic gastritis-associated dyspepsia"
gene <1. .>291
/gene="cagA"
CDS <1. .>291
/gene="cagA"
/note="CagA"
/codon_start=1
/transl_table=11

/product="cytotoxin associated protein A"
/protein_id="AAG09869.1"
/db_xref="GI:9964195"
/translation="IKNPTKKNQYFSDFINKSNDLINKDALIDVKSSTKSFQKFGDQR YRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFLKSAKQSFAG"
misc_feature 115. .135
/gene="cagA"
/note="Region: tyrosine phosphorylation motif A"
BASE COUNT 109 a 62 c 43 g 77 t
ORIGIN

Query Match 2.4%; Score 144; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 4.3e-57;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    842 aatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagtt 901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 AATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTT 151

Qy    902 gggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatgg 961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 GGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGG 211

Qy    962 aaaatatcatacaaccccctatcc 985
          |||||||||||||||||||||||
Db    212 AAAATATCATACAACCCCCTATCC 235
```

RESULT 49
AF289400
LOCUS AF289400 291 bp DNA BCT 02-SEP-2000
DEFINITION Helicobacter pylori isolate CR2 cytotoxin associated protein A (cagA) gene, partial cds.
ACCESSION AF289400
VERSION AF289400.1 GI:9964154
KEYWORDS .
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 291)
AUTHORS Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N., Covacci,A. and Megraud,F.
TITLE Composition and gene expression of the cag pathogenicity island in Helicobacter pylori strains isolated from gastric carcinoma and gastritis patients in Costa Rica
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 291)
AUTHORS Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N., Covacci,A. and Megraud,F.

TITLE Direct Submission
JOURNAL Submitted (26-JUL-2000) Laboratory of Bacteriology, University of Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES Location/Qualifiers
source 1. .291
/organism="Helicobacter pylori"
/isolate="CR2"
/db_xref="taxon:210"
/country="Costa Rica"
/note="isolated from a patient with gastric adenocarcinoma"
gene <1. .>291
/gene="cagA"
CDS <1. .>291
/gene="cagA"
/note="CagA"
/codon_start=1
/transl_table=11
/product="cytotoxin associated protein A"
/protein_id="AAG09849.1"
/db_xref="GI:9964155"
/translation="IKNPTKKNQYFSDFINKSNDLINKDALIDVEPSTKSFQKFGDQR YRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFLKSAKQSFAG"
misc_feature 115. .135
/gene="cagA"
/note="Region: tyrosine phosphorylation motif A"
BASE COUNT 108 a 63 c 44 g 76 t
ORIGIN

Query Match 2.4%; Score 141; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.2e-55;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    845 cttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagttggg 904
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     95 CTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGG 154

Qy    905 tgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaa 964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    155 TGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAAA 214

Qy    965 atatcatacaaccccctatcc 985
          |||||||||||||||||||||
Db    215 ATATCATACAACCCCCTATCC 235
```

RESULT 50
AF198482
LOCUS AF198482 220 bp DNA BCT 19-MAY-2000
DEFINITION Helicobacter pylori strain PeruGJ3 cytotoxin associated protein

CagA (cagA) gene, partial cds.
ACCESSION AF198482
VERSION AF198482.1 GI:7248690
KEYWORDS .
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 220)
AUTHORS van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M., Xiao,S.D., Tytgat,G.N. and Dankert,J.
TITLE cagA-positive Helicobacter pylori populations in China and The Netherlands are distinct
JOURNAL Infect. Immun. 66 (5), 1822-1826 (1998)
MEDLINE 98234000
PUBMED 9573056
REFERENCE 2 (bases 1 to 220)
AUTHORS Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z., Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H., Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G., Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I., Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
TITLE Differences in genotypes of Helicobacter pylori from different human populations
JOURNAL J. Bacteriol. 182 (11), 3210-3218 (2000)
MEDLINE 20270152
PUBMED 10809702
REFERENCE 3 (bases 1 to 220)
AUTHORS Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J. and Berg,D.E.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1999) Molecular Microbiology, Washington University Medical School, 4566 Scott Ave., St. Louis, MO 63110, USA
FEATURES Location/Qualifiers
source 1. .220
/organism="Helicobacter pylori"
/strain="PeruGJ3"
/db_xref="taxon:210"
/note="European type; isolated from gastric juice sample from native Peruvian"
gene <1. .>220
/gene="cagA"
CDS <1. .>220
/gene="cagA"
/note="similar to Helicobacter pylori strain 26695 HP0547"
/codon_start=2
/transl_table=11
/product="cytotoxin associated protein CagA"
/protein_id="AAF43622.1"

```
                     /db_xref="GI:7248691"
                     /translation="KDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSPYLYKEDQL
                     TGSQRALSQEEIQNKIDFMEFLAQNNAKL"
BASE COUNT     89 a     39 c     42 g     50 t
ORIGIN


  Query Match             2.4%;  Score 140;  DB 1;  Length 220;
  Best Local Similarity   99.5%;  Pred. No. 3.5e-55;
  Matches  190;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     17 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 76

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 136

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    137 GGCTCACAACGAGCATTGAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 196

Qy   1741 cttgcacaaaa 1751
          ||||||||||| 
Db    197 CTTGCACAAAA 207


RESULT  51
HPY239736
LOCUS       HPY239736     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain 5596.
ACCESSION   AJ239736
VERSION     AJ239736.1  GI:4467790
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
```



```
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1..243
                     /organism="Helicobacter pylori"
                     /strain="5596"
                     /db_xref="taxon:210"
                     /country="Gambia"
     gene            1..243
                     /gene="cagA"
     CDS             <1..>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37828.1"
                     /db_xref="GI:4467791"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSPYLYKE
                     DQLTGSQRALSQEEIQNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       95 a     43 c     50 g     55 t
ORIGIN


  Query Match             2.4%;  Score 140;  DB 2;  Length 243;
  Best Local Similarity   99.5%;  Pred. No. 3.5e-55;
  Matches  190;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    145 GGCTCACAACGAGCATTGAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 204

Qy   1741 cttgcacaaaa 1751
          ||||||||||| 
Db    205 CTTGCACAAAA 215


RESULT  52
AF198468
LOCUS       AF198468      281 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain Gambia4659 cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198468
VERSION     AF198468.1  GI:7248662
KEYWORDS
```



```
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 281)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 281)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1..281
                     /organism="Helicobacter pylori"
                     /strain="Gambia4659"
                     /db_xref="taxon:210"
                     /note="European type"
     gene            <1..>281
                     /gene="cagA"
     CDS             <1..>281
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43608.1"
                     /db_xref="GI:7248663"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSPYLYKEDQLTGSQRALSQEEIQNKIDFMEFLAQNNAKLDNL"
BASE COUNT      103 a     53 c     62 g     63 t
ORIGIN
```



```
  Query Match             2.4%;  Score 140;  DB 1;  Length 281;
  Best Local Similarity   99.5%;  Pred. No. 3.5e-55;
  Matches  190;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 187

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    188 GGCTCACAACGAGCATTGAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 247

Qy   1741 cttgcacaaaa 1751
          ||||||||||| 
Db    248 CTTGCACAAAA 258


RESULT  53
AF198477
LOCUS       AF198477      290 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain Peru4A cytotoxin associated protein CagA
            (cagA) gene, partial cds.
ACCESSION   AF198477
VERSION     AF198477.1  GI:7248680
KEYWORDS
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 290)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 290)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
```

```
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 290)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .290
                     /organism="Helicobacter pylori"
                     /strain="Peru4A"
                     /db_xref="taxon:210"
                     /country="Peru"
                     /note="European type"
     gene            <1. .>290
                     /gene="cagA"
     CDS             <1. .>290
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43617.1"
                     /db_xref="GI:7248681"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGSQRALSQEEIQNKIDFMEFLAQNNAKLDNLSEK"
BASE COUNT      108 a      52 c      66 g      64 t
ORIGIN
```

```
  Query Match             2.4%;  Score 140;  DB 1;  Length 290;
  Best Local Similarity   99.5%;  Pred. No. 3.5e-55;
  Matches  190;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 187

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    188 GGCTCACAACGAGCATTGAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 247

Qy   1741 cttgcacaaaa 1751
          ||||||||||| 
Db    248 CTTGCACAAAA 258
```



```
RESULT  54
HPY239690
LOCUS       HPY239690     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain 25.
ACCESSION   AJ239690
VERSION     AJ239690.1  GI:4467698
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Helicobacter pylori"
                     /strain="25"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .413
                     /gene="cagA"
     CDS             <1. .>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37801.1"
                     /db_xref="GI:4467699"
                     /translation="EAAPNPQQPINNLQVAPLKVDNAVASYDPDQKPIVDKNDRDNRQ
                     AFDGISKLREEYSNKAIKNPTKKNQYPSDPIDKSNDLINKDNLIDVESSTKSPQKFGD
                     QRYRIFTSWVSHQNDPGKINTRSIRNFMENIIQPP"
BASE COUNT      151 a      81 c      68 g     113 t
ORIGIN
```

```
  Query Match             2.4%;  Score 140;  DB 2;  Length 413;
```



```
  Best Local Similarity   100.0%;  Pred. No. 3.4e-55;
  Matches  140;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    571 gaagcggcttttaacccgcagcaatttatcaataatcttcaagtagcttttcttaaagtt 630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGTT 60

Qy    631 gataacgctgtcgcttcatacgatcctgatcaaaaaccaatcgttgataagaacgatagg 690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGG 120

Qy    691 gataacaggcaagcttttga 710
          ||||||||||||||||||||
Db    121 GATAACAGGCAAGCTTTTGA 140


RESULT  55
HPY269884
LOCUS       HPY269884     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            US178 from United States.
ACCESSION   AJ269884
VERSION     AJ269884.1  GI:5921345
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="US178"
                     /db_xref="taxon:210"
                     /country="USA"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
```



```
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56490.1"
                     /db_xref="GI:5921346"
                     /translation="MTNETIDQQPQTEAAFNPQRFINNLQVAPLKLDNAVASFDPDQK
                     PIVDKNDRDNRQAFEGISQLREEYSNKAIKNPTKKNQYFSDPINKSNDLINKDNLIDV
                     ESSAKSFQKFGDQRYRIFTSWVSHQKDPSKINTRSIRNFMEHTIQPP"
BASE COUNT      163 a      98 c      73 g     115 t
ORIGIN
```

```
  Query Match             2.3%;  Score 138;  DB 2;  Length 449;
  Best Local Similarity   99.5%;  Pred. No. 3e-54;
  Matches  188;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    651 cgatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttga 710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    117 CGATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGA 176

Qy    711 aggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaa 770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 GGGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAA 236

Qy    771 gaatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatct 830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 GAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCT 296

Qy    831 cattgatgt 839
          |||||||||
Db    297 CATTGATGT 305


RESULT  56
AB003397
LOCUS       AB003397     3744 bp    DNA             BCT       05-FEB-1999
DEFINITION  Helicobacter pylori DNA for CagA, complete cds.
ACCESSION   AB003397
VERSION     AB003397.1  GI:2073134
KEYWORDS    CagA.
SOURCE      Helicobacter pylori (strain:ATCC43526) DNA.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 3744)
  AUTHORS   Maeda,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-1997) to the DDBJ/EMBL/GenBank databases. Shin
            Maeda, University of Tokyo, Second Department of Internal Medicine;
            7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
            (E-mail:MAEDA-2IM@h.u-tokyo.ac.jp, Tel:81-3-3815-5411)
REFERENCE   2  (sites)
```

```
  AUTHORS   Maeda,S., Kanai,F., Ogura,K., Yoshida,H., Ikenoue,T., Kawabe,T.,
            Shiratori,Y. and Omata,M.
  TITLE     High Prevalence of CagA-Positive Helicobacter pylori in Japanese
            Patien ts Irrelevant to the Presence of Peptic Ulcers and Gastric
            Cancer
  JOURNAL   Unpublished (1997)
FEATURES             Location/Qualifiers
     source          1. .3744
                     /organism="Helicobacter pylori"
                     /strain="ATCC43526"
                     /db_xref="taxon:210"
     gene            1. .3744
                     /gene="cagA"
     CDS             1. .3744
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA"
                     /protein_id="BAA19853.1"
                     /db_xref="GI:2073135"
                     /translation="MTNETIAQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDI
                     GSSIKSPQKFGTQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPTPDDKERAEFV
                     KSAKQSFAGIIIGNQIRTDQKFMGVFDEFLKERQEAEKNGEPTGGDWLDIFLSFVFNK
                     EQSSDVKEAINQEPVPHVQPDIATTTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLD
                     VEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLFAGNGGFGAKHDWNAT
                     VGYKNQQGDNVATLINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQLTGSQRALSQEE
                     IRNKIDFMEFLAQNNAKLDNLSEKEKEKFQNEIEDFQKDSKAYLDALGNDRIAFVSKK
                     DPKHPALITEFGKGDLSYTLKVMGKKQTEALDREKNVTLQGNLKHDSVMFVNYSNFKY
                     TNASKSPDKGVGVTNGVSHLDAGFSKVAVFNLPDLNNLAITSFVRRNLENKLVTEGLS
                     LQEANKLIKDFLSSNKELVGKALNFNKAVADAKNTGDYDEVKKAQKDLEKSLRKREHL
                     EKEVEKKLETKSGNKNKMEAKAQANGQKDKIFALINKEANRDARAIAYSQNLKGIKRE
                     LSDKLEKINTDLKDFSKSFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVE
                     NLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKITDKVDNLNQAVSMAKATG
                     GFSRVEQALADLKNFSKEQLAQQTQKNESFNVGKKSEIYQSVKNGVNGTLVGNGLSGI
                     EATALAKNFSDIKKELNEKFKNFNNNNNNGLENEPIYAKVNKKKTGQVASPEEPIYAQ
                     VAKKVNAKIDRLNQAASGLGGVGQAGFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLG
                     GPFPSKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPE
                     PIYATIDDLGGPFPSKRHDKVDDLSKVGLSRNQELAQKIDNLSQAVSEAKAGFFSNLE
                     QTIDKLKDSTKYNSVNLWVESAQKVPASLSAKLDNYATNSHTRINSNIQNGAINEKAT
                     GMLTHKNPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMKDYSDSFKFSTKLNNAV
                     KDVKSSFTQFLANAFSTGYYSLARENAEHGIKNVNTKGGFQKS"
BASE COUNT     1351 a     667 c     777 g     949 t
ORIGIN
```

```
  Query Match             2.3%;  Score 138;  DB 1;  Length 3744;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-54;
  Matches  138;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2734 gaatggatttcaaaagttgaaaaccttaatgcagctttgaatgaattcaaaaatggcaaa 2793
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2200 GAATGGATTTCAAAAGTTGAAAACCTTAATGCAGCTTTGAATGAATTCAAAAATGGCAAA 2259

Qy   2794 aataaggatttcagcaaggtaacgcaagcaaaaagcgaccttgaaaattccgttaaagat 2853
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2260 AATAAGGATTTCAGCAAGGTAACGCAAGCAAAAAGCGACCTTGAAAATTCCGTTAAAGAT 2319

Qy   2854 gtgatcatcaatcaaaag 2871
          ||||||||||||||||||
Db   2320 GTGATCATCAATCAAAAG 2337
```

```
RESULT  57
AF001357
LOCUS       AF001357     3744 bp    DNA             BCT       03-JUN-1997
DEFINITION  Helicobacter pylori strain ATCC 43526 CagA (cagA) gene, complete
            cds.
ACCESSION   AF001357
VERSION     AF001357.1  GI:2150121
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 3744)
  AUTHORS   Maeda,S., Kanai,F., Ogura,K., Yoshida,H., Ikenoue,T., Kawabe,T.,
            Shiratori,Y. and Omata,M.
  TITLE     High Seropositivity of Anti-CagA Antibody in Helicobacter pylori
            Infected Patienta Irrelevant to Peptic Ulcers and Normal Mucosa in
            Japan
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3744)
  AUTHORS   Maeda,S., Kanai,F., Ogura,K., Yoshida,H., Ikenoue,T., Kawabe,T.,
            Shiratori,Y. and Omata,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-1997) Second Department of Internal Medicine,
            University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
FEATURES             Location/Qualifiers
     source          1. .3744
                     /organism="Helicobacter pylori"
                     /strain="ATCC 43526"
                     /db_xref="taxon:210"
     gene            1. .3744
                     /gene="cagA"
     CDS             1. .3744
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA"
                     /protein_id="AAB58747.1"
                     /db_xref="GI:2150122"
```



```
                     /translation="MTNETIAQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDI
                     GSSIKSFQKFGTQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPTPDDKERAEFV
                     KSAKQSFAGIIIGNQIRTDQKFMGVFDEFLKERQEAEKNGEPTGGDWLDIFLSFVFNK
                     EQSSDVKEAINQEPVPHVQPDIATTTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLD
                     VEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLFAGNGGFGAKHDWNAT
                     VGYKNQQGDNVATLINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQLTGSQRALSQEE
                     IRNKIDFMEFLAQNNAKLDNLSEKEKEKFQNEIEDFQKDSKAYLDALGNDRIAFVSKK
                     DPKHPALITEFGKGDLSYTLKVMGKKQTEALDREKNVTLQGNLKHDSVMFVNYSNFKY
                     TNASKSPDKGVGVTNGVSHLDAGFSKVAVFNLPDLNNLAITSFVRRNLENKLVTEGLS
                     LQEANKLIKDFLSSNKELVGKALNFNKAVADAKNTGDYDEVKKAQKDLEKSLRKREHL
                     EKEVEKKLETKSGNKNKMEAKAQANGQKDKIFALINKEANRDARAIAYSQNLKGIKRE
                     LSDKLEKINTDLKDFSKSFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVE
                     NLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKITDKVDNLNQAVSMAKATG
                     GFSRVEQALADLKNFSKEQLAQQTQKNESFNVGKKSEIYQSVKNGVNGTLVGNGLSGI
                     EATALAKNFSDIKKELNEKFKNFNNNNNNGLENEPIYAKVNKKKTGQVASPEEPIYAQ
                     VAKKVNAKIDRLNQAASGLGGVGQAGFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLG
                     GPFPSKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPE
                     PIYATIDDLGGPFPSKRHDKVDDLSKVGLSRNQELAQKIDNLSQAVSEAKAGFFSNLE
                     QTIDKLKDSTKYNSVNLWVESAQKVPASLSAKLDNYATNSHTRINSNIQNGAINEKAT
                     GMLTHKNPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMKDYSDSFKFSTKLNNAV
                     KDVKSSFTQFLANAFSTGYYSLARENAEHGIKNVNTKGGFQKS"
BASE COUNT     1351 a     667 c     777 g     949 t
ORIGIN
```

```
  Query Match             2.3%;  Score 138;  DB 1;  Length 3744;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-54;
  Matches  138;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2734 gaatggatttcaaaagttgaaaaccttaatgcagctttgaatgaattcaaaaatggcaaa 2793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2200 GAATGGATTTCAAAAGTTGAAAACCTTAATGCAGCTTTGAATGAATTCAAAAATGGCAAA 2259

Qy   2794 aataaggatttcagcaaggtaacgcaagcaaaaagcgaccttgaaaattccgttaaagat 2853
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2260 AATAAGGATTTCAGCAAGGTAACGCAAGCAAAAAGCGACCTTGAAAATTCCGTTAAAGAT 2319

Qy   2854 gtgatcatcaatcaaaag 2871
          ||||||||||||||||||
Db   2320 GTGATCATCAATCAAAAG 2337
```

```
RESULT  58
AF202973
LOCUS       AF202973     3744 bp    DNA             BCT       02-MAR-2000
DEFINITION  Helicobacter pylori strain NCTC 11637 CagA (cagA) gene, complete
            cds.
ACCESSION   AF202973
VERSION     AF202973.1  GI:6573221
KEYWORDS    .
```



```
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 3744)
  AUTHORS   Asahi,M., Azuma,T., Ito,S., Ito,Y., Suto,H., Nagai,Y.,
            Tsubokawa,M., Tohyama,Y., Maeda,S., Omata,M., Suzuki,T. and
            Sasakawa,C.
  TITLE     Helicobacter pylori CagA protein can be tyrosine phosphorylated in
            gastric epithelial cells
  JOURNAL   J. Exp. Med. 191 (4), 593-602 (2000)
  MEDLINE   20150113
REFERENCE   2  (bases 1 to 3744)
  AUTHORS   Ito,Y. and Azuma,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-1999) Second Department of Internal Medicine,
            Fukui Medical University, Matsuoka-cho, Yoshida-gun, Fukui
            910-1193, Japan
FEATURES             Location/Qualifiers
     source          1. .3744
                     /organism="Helicobacter pylori"
                     /strain="NCTC 11637"
                     /db_xref="taxon:210"
     gene            1. .3744
                     /gene="cagA"
     CDS             1. .3744
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA"
                     /protein_id="AAF17598.1"
                     /db_xref="GI:6573222"
                     /translation="MTNETIDQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDI
                     GSSIKSFQKFGTQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFL
                     KSAKQSFAGIIIGNQIRTDQKFMGVFDEFLKERQEAEKNGEPTGGDWLDIFLSFVFNK
                     EQSSDVKEAINQEPVPHVQPDIATTTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLD
                     VEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNAT
                     VGYKNQQGDNVATLINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQLTGSQRALSQEE
                     IRNKIDFMEFLAQNNAKLDNLSEKEKEKFQNEIEDFQKDSKAYLDALGNDRIAFVSKK
                     DPKHSALITEFGKGDLSYTLKDYGKKADRALDREKNVTLQGNLKHDSVMFVNYSNFKY
                     TNASKSPDKGVGVTNGVSHLDAGFSKVAVFNLPDLNNLAITSFVRRNLENKLVTEGLS
                     LQEANKLIKDFLSSNKELVGKALNFNKAVADAKNTGNYDEVKKAQKDLEKSLRKREHL
                     EKEVEKKLESKSGNKNKMEAKAQANSQKDKIFALINKEANRDARAIAYSQNLKGIKRE
                     LSDKLEKINKDLKDFSKSFDEFKNGKNKDFSKAEETLKALKGSVRDLGINPEWISKVE
                     NLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKITDKVDNLNQAVSMAKATG
                     DFSRVEQALADLKNFSKEQLAQQTQKNESFNVGKKSEIYQSVKNGVNGTLVGNGLSGI
                     EATALAKNFSDIKKELNEKFKNFNNNNNNGLENEPIYAKVNKKKTGQVASPEEPIYAQ
                     VAKKVNAKIDRLNQAASGLGGVGQAGFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLG
                     GPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPE
                     PIYATIDDLGGPFPLKRHDKVDDLSKVGLSRNQELAQKIDNLSQAVSEAKAGFFSNLE
```

QTIDKLKDSTKYNSVNLWVESAKKVPASLSAKLDNYATNSHTRINSNIQNGAINEKAT
GMLTQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMKDYSDSFKFSTKLNNAV
KDVKSSFTQFLANAFSTGYYSLAREHAEHGIKNVNTKGGFQKS"
BASE COUNT 1363 a 657 c 769 g 955 t
ORIGIN

Query Match 2.3%; Score 138; DB 1; Length 3744;
Best Local Similarity 100.0%; Pred. No. 2.6e-54;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2734 gaatggatttcaaaagttgaaaaccttaatgcagctttgaatgaattcaaaaatggcaaa 2793
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2200 GAATGGATTTCAAAAGTTGAAAACCTTAATGCAGCTTTGAATGAATTCAAAAATGGCAAA 2259

Qy 2794 aataaggatttcagcaaggtaacgcaagcaaaaagcgaccttgaaaattccgttaaagat 2853
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2260 AATAAGGATTTCAGCAAGGTAACGCAAGCAAAAAGCGACCTTGAAAATTCCGTTAAAGAT 2319

Qy 2854 gtgatcatcaatcaaaag 2871
|||||||||||||||||||
Db 2320 GTGATCATCAATCAAAAG 2337

RESULT 59
AF198474
LOCUS AF198474 220 bp DNA BCT 19-MAY-2000
DEFINITION Helicobacter pylori strain Peru002B cytotoxin associated protein CagA (cagA) gene, partial cds.
ACCESSION AF198474
VERSION AF198474.1 GI:7248674
KEYWORDS .
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 220)
AUTHORS van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M., Xiao,S.D., Tytgat,G.N. and Dankert,J.
TITLE cagA-positive Helicobacter pylori populations in China and The Netherlands are distinct
JOURNAL Infect. Immun. 66 (5), 1822-1826 (1998)
MEDLINE 98234000
PUBMED 9573056
REFERENCE 2 (bases 1 to 220)
AUTHORS Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z., Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H., Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G., Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I., Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.

TITLE Differences in genotypes of Helicobacter pylori from different human populations
JOURNAL J. Bacteriol. 182 (11), 3210-3218 (2000)
MEDLINE 20270152
PUBMED 10809702
REFERENCE 3 (bases 1 to 220)
AUTHORS Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J. and Berg,D.E.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1999) Molecular Microbiology, Washington University Medical School, 4566 Scott Ave., St. Louis, MO 63110, USA
FEATURES Location/Qualifiers
source 1..220
/organism="Helicobacter pylori"
/strain="Peru002B"
/db_xref="taxon:210"
/country="Peru"
/note="European type"
gene <1..>220
/gene="cagA"
CDS <1..>220
/gene="cagA"
/note="similar to Helicobacter pylori strain 26695 HP0547"
/codon_start=2
/transl_table=11
/product="cytotoxin associated protein CagA"
/protein_id="AAF43614.1"
/db_xref="GI:7248675"
/translation="KDLQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQL TGSQRALSQEEIRNKIDFMEFLAQNNAKL"
BASE COUNT 86 a 39 c 43 g 52 t
ORIGIN

Query Match 2.3%; Score 137; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 9.4e-54;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 17 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 76

Qy 1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 77 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 136

Qy 1681 ggctcacaacgagcatt 1697
|||||||||||||||||
Db 137 GGCTCACAACGAGCATT 153

RESULT 60
HPY239732
LOCUS HPY239732 243 bp DNA BCT 30-JUN-1999
DEFINITION Helicobacter pylori partial cagA gene, strain 25.
ACCESSION AJ239732
VERSION AJ239732.1 GI:4467782
KEYWORDS cagA gene; cytotoxin associated protein A.
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 243)
AUTHORS Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J., Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
TITLE Recombination and clonal groupings within helicobacter pylori from different geographical regions
JOURNAL Mol. Microbiol. 32 (3), 459-470 (1999)
MEDLINE 99255683
REFERENCE 2 (bases 1 to 243)
AUTHORS van der Ende,A.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1999) van der Ende A., Medical Microbiology, Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ, NETHERLANDS
FEATURES Location/Qualifiers
source 1..243
/organism="Helicobacter pylori"
/strain="25"
/db_xref="taxon:210"
/country="Netherlands"
gene 1..243
/gene="cagA"
CDS <1..>243
/gene="cagA"
/codon_start=1
/transl_table=11
/product="cytotoxin associated protein A"
/protein_id="CAB37824.1"
/db_xref="GI:4467783"
/db_xref="SPTREMBL:Q9Z5L6"
/translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE DQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSV"
BASE COUNT 93 a 42 c 51 g 57 t
ORIGIN

Query Match 2.3%; Score 137; DB 2; Length 243;
Best Local Similarity 100.0%; Pred. No. 9.4e-54;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy 1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 85 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy 1681 ggctcacaacgagcatt 1697
|||||||||||||||||
Db 145 GGCTCACAACGAGCATT 161

RESULT 61
HPY239737
LOCUS HPY239737 243 bp DNA BCT 30-JUN-1999
DEFINITION Helicobacter pylori partial cagA gene, strain CC28.
ACCESSION AJ239737
VERSION AJ239737.1 GI:4467792
KEYWORDS cagA gene; cytotoxin associated protein A.
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 243)
AUTHORS Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J., Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
TITLE Recombination and clonal groupings within helicobacter pylori from different geographical regions
JOURNAL Mol. Microbiol. 32 (3), 459-470 (1999)
MEDLINE 99255683
REFERENCE 2 (bases 1 to 243)
AUTHORS van der Ende,A.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1999) van der Ende A., Medical Microbiology, Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ, NETHERLANDS
FEATURES Location/Qualifiers
source 1..243
/organism="Helicobacter pylori"
/strain="CC28"
/db_xref="taxon:210"
/country="South Africa"
gene 1..243
/gene="cagA"
CDS <1..>243
/gene="cagA"
/codon_start=1
/transl_table=11
/product="cytotoxin associated protein A"
/protein_id="CAB37829.1"

```
                     /db_xref="GI:4467793"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIQNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       95 a      41 c      50 g      57 t
ORIGIN


  Query Match             2.3%;  Score 137;  DB 2;  Length 243;
  Best Local Similarity   99.5%;  Pred. No. 9.4e-54;
  Matches  187;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1564 aatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagcaggt 1623
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     28 AATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCAGGT 87

Qy   1624 ggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcacaggc 1683
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     88 GGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACAGGC 147

Qy   1684 tcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaatttctt 1743
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    148 TCACAACGAGCATTGAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTTCTT 207

Qy   1744 gcacaaaa 1751
          ||||||||
Db    208 GCACAAAA 215


RESULT  62
HPY252968
LOCUS       HPY252968     244 bp    DNA             BCT       08-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
            protein A.
ACCESSION   AJ252968
VERSION     AJ252968.1  GI:5851967
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
REFERENCE   2  (bases 1 to 244)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) van der Ende A.; Medical Microbiology,
```



```
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Helicobacter pylori"
                     /strain="Dutch 25"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .244
                     /gene="cagA"
     CDS             <1. .>244
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB55444.1"
                     /db_xref="GI:5851968"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSV"
BASE COUNT       93 a      42 c      52 g      57 t
ORIGIN


  Query Match             2.3%;  Score 137;  DB 2;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-54;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    145 GGCTCACAACGAGCATT 161


RESULT  63
HPY252969
LOCUS       HPY252969     244 bp    DNA             BCT       08-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
            protein A.
ACCESSION   AJ252969
VERSION     AJ252969.1  GI:5851969
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
```



```
            Helicobacter.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
REFERENCE   2  (bases 1 to 244)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Helicobacter pylori"
                     /strain="Dutch 27"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .244
                     /gene="cagA"
     CDS             <1. .>244
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB55445.1"
                     /db_xref="GI:5851970"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       94 a      42 c      52 g      56 t
ORIGIN


  Query Match             2.3%;  Score 137;  DB 2;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-54;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    145 GGCTCACAACGAGCATT 161
```



```
RESULT  64
HPY252971
LOCUS       HPY252971     244 bp    DNA             BCT       08-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
            protein A.
ACCESSION   AJ252971
VERSION     AJ252971.1  GI:5851973
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
REFERENCE   2  (bases 1 to 244)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Helicobacter pylori"
                     /strain="Dutch 292"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .244
                     /gene="cagA"
     CDS             <1. .>244
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB55447.1"
                     /db_xref="GI:5851974"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       95 a      41 c      51 g      57 t
ORIGIN


  Query Match             2.3%;  Score 137;  DB 2;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-54;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    145 GGCTCACAACGAGCATT 161
```

RESULT 65
HPY252974
LOCUS HPY252974 244 bp DNA BCT 08-SEP-1999
DEFINITION Helicobacter pylori partial cagA gene for cytotoxin associated protein A.
ACCESSION AJ252974
VERSION AJ252974.1 GI:5851979
KEYWORDS cagA gene; cytotoxin associated protein A.
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 244)
AUTHORS van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M., Xiao,S.D., Tytgat,G.N. and Dankert,J.
TITLE cagA-positive Helicobacter pylori populations in China and The Netherlands are distinct
JOURNAL Infect. Immun. 66 (5), 1822-1826 (1998)
MEDLINE 98234000
REFERENCE 2 (bases 1 to 244)
AUTHORS van der Ende,A.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-1999) van der Ende A., Medical Microbiology, Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ, NETHERLANDS
FEATURES Location/Qualifiers
source 1..244
/organism="Helicobacter pylori"
/strain="Dutch 419"
/db_xref="taxon:210"
/country="Netherlands"
gene 1..244
/gene="cagA"
CDS <1..>244
/gene="cagA"
/codon_start=1
/transl_table=11
/product="cytotoxin associated protein A"

/protein_id="CAB55450.1"
/db_xref="GI:5851980"
/translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE DQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT 95 a 42 c 51 g 56 t
ORIGIN

```
Query Match           2.3%;  Score 137;  DB 2;  Length 244;
Best Local Similarity  100.0%;  Pred. No. 9.4e-54;
Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaatattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    145 GGCTCACAACGAGCATT 161
```

RESULT 66
AF198469
LOCUS AF198469 281 bp DNA BCT 19-MAY-2000
DEFINITION Helicobacter pylori strain Gambia4797 cytotoxin associated protein CagA (cagA) gene, partial cds.
ACCESSION AF198469
VERSION AF198469.1 GI:7248664
KEYWORDS .
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 281)
AUTHORS van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M., Xiao,S.D., Tytgat,G.N. and Dankert,J.
TITLE cagA-positive Helicobacter pylori populations in China and The Netherlands are distinct
JOURNAL Infect. Immun. 66 (5), 1822-1826 (1998)
MEDLINE 98234000
PUBMED 9573056
REFERENCE 2 (bases 1 to 281)
AUTHORS Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z., Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H., Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G., Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I., Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,

Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
TITLE Differences in genotypes of Helicobacter pylori from different human populations
JOURNAL J. Bacteriol. 182 (11), 3210-3218 (2000)
MEDLINE 20270152
PUBMED 10809702
REFERENCE 3 (bases 1 to 281)
AUTHORS Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J. and Berg,D.E.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1999) Molecular Microbiology, Washington University Medical School, 4566 Scott Ave., St. Louis, MO 63110, USA
FEATURES Location/Qualifiers
source 1..281
/organism="Helicobacter pylori"
/strain="Gambia4797"
/db_xref="taxon:210"
/country="Gambia"
/note="European type"
gene <1..>281
/gene="cagA"
CDS <1..>281
/gene="cagA"
/note="similar to Helicobacter pylori strain 26695 HP0547"
/codon_start=2
/transl_table=11
/product="cytotoxin associated protein CagA"
/protein_id="AAF43609.1"
/db_xref="GI:7248665"
/translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG EKGINNPSFYLYKEDQLTGSQRALSQEEILNKIDFMEFLAQNNAKLDNL"
BASE COUNT 102 a 52 c 62 g 65 t
ORIGIN

```
Query Match           2.3%;  Score 137;  DB 1;  Length 281;
Best Local Similarity  100.0%;  Pred. No. 9.3e-54;
Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 187

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    188 GGCTCACAACGAGCATT 204
```


RESULT 67
AF202221
LOCUS AF202221 287 bp DNA BCT 19-MAY-2000
DEFINITION Helicobacter pylori strain India9A CagA (cagA) gene, partial cds.
ACCESSION AF202221
VERSION AF202221.1 GI:7229433
KEYWORDS .
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
REFERENCE 1 (bases 1 to 287)
AUTHORS Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y., Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T., Nair,G.B. and Berg,D.E.
TITLE Distinctiveness of genotypes of Helicobacter pylori in Calcutta, India
JOURNAL J. Bacteriol. 182 (11), 3219-3227 (2000)
MEDLINE 20270153
PUBMED 10809703
REFERENCE 2 (bases 1 to 287)
AUTHORS Mukhopadhyay,A.K., Kersulyte,D., Chowdhury,A., Datta,S., Nair,G.B. and Berg,D.E.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) Molecular Microbiology, Washington University, 660 South Euclid Ave., St. Louis, MO 63110, USA
FEATURES Location/Qualifiers
source 1..287
/organism="Helicobacter pylori"
/strain="India9A"
/db_xref="taxon:210"
gene <1..>287
/gene="cagA"
CDS <1..>287
/gene="cagA"
/note="similar to Helicobacter pylori 26695 HP0547"
/codon_start=2
/transl_table=11
/product="CagA"
/protein_id="AAF42823.1"
/db_xref="GI:7229434"
/translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT 104 a 52 c 66 g 65 t
ORIGIN

```
Query Match           2.3%;  Score 137;  DB 1;  Length 287;
Best Local Similarity  100.0%;  Pred. No. 9.3e-54;
```

```
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 187

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    188 GGCTCACAACGAGCATT 204


RESULT  68
AF198471
LOCUS       AF198471      299 bp    DNA              BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain SAfrica56 cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198471
VERSION     AF198471.1  GI:7248668
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populationa in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 299)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 299)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Helicobacter pylori"
                     /strain="SAfrica56"
                     /db_xref="taxon:210"
                     /country="South Africa"
                     /note="European type"
     gene            <1. .>299
                     /gene="cagA"
     CDS             <1. .>299
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=3
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43611.1"
                     /db_xref="GI:7248669"
                     /translation="ANGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSPYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSEKEKE"
BASE COUNT      113 a     53 c     69 g     64 t
ORIGIN


  Query Match             2.3%;  Score 137;  DB 1;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-54;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 128

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 188

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    189 GGCTCACAACGAGCATT 205


RESULT  69
HPY269880
LOCUS       HPY269880     449 bp    DNA              BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            PO16 from Portugal.
ACCESSION   AJ269880
VERSION     AJ269880.1  GI:5921337
KEYWORDS    cagA gene; CagA protein.
```



```
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="PO16"
                     /db_xref="taxon:210"
                     /country="Portugal"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56448.1"
                     /db_xref="GI:5921338"
                     /translation="MTNETIDQQPQTEAAFNPQQFINNLQVAFLKLDNAVASFDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDALIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIQNFMENIIQPP"
BASE COUNT      163 a     95 c     71 g     120 t
ORIGIN


  Query Match             2.3%;  Score 136;  DB 2;  Length 449;
  Best Local Similarity   99.2%;  Pred. No. 2.7e-53;
  Matches  236;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaaag 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 GGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAG 237

Qy    772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db    238 AATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACGCTCTC 297

Qy    832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 ATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATT 357

Qy    892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatcc 949
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 TTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCC 415


RESULT  70
AF289402
LOCUS       AF289402      291 bp    DNA              BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR5 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289402
VERSION     AF289402.1  GI:9964158
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR5"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09851.1"
                     /db_xref="GI:9964159"
```

```
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSAKSFQKFGDQR
                     YQIFTSWVSHQNDPSKINTRSIRNFMENIIQPPISDDKEKAEFLKSAKQSFAG"
     misc_feature    115..135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      108 a     59 c     45 g     79 t
ORIGIN
```

```
  Query Match             2.2%;  Score 132;  DB 2;  Length 291;
  Best Local Similarity   99.1%;  Pred. No. 2.2e-51;
  Matches  232;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 60

Qy    811 ttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaattt 870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCGCAAAGAGCTTTCAGAAATTT 120

Qy    871 ggggatcagcgttaccgaatttttcacaagttgggtgtcccatcaaaacgatccgtctaaa 930
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGGATCAGCGTTACCAAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAA 180

Qy    931 atcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccccctatc 984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATACAACCCCCTATC 234
```

```
RESULT  71
HPY269886
LOCUS       HPY269886     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate AU1
            from Australia.
ACCESSION   AJ269886
VERSION     AJ269886.1  GI:5921349
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1..449
                     /organism="Helicobacter pylori"
                     /isolate="AU1"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1..449
                     /gene="cagA"
     CDS             1..>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56451.1"
                     /db_xref="GI:5921350"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQR
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDI
                     GSSIKSFQKFGTQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      167 a     96 c     68 g    118 t
ORIGIN
```

```
  Query Match             2.2%;  Score 131;  DB 2;  Length 449;
  Best Local Similarity   98.9%;  Pred. No. 6.5e-51;
  Matches  281;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggctttaacccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 ACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAG 79

Qy    614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 TGGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCG 139

Qy    674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
          |||||||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    140 TTGATAAGAATGATAGGGATAACAGGCAAGCTTTTGATGGAATCTCGCAATTAAGGGAAG 199

Qy    734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 AATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTA 259

Qy    794 tcaataagagcaatgatttaatcaacaaagacaatctcattgat 837
          |||||||||||||||||||||||||||||||||||||||||||
Db    260 TCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGAT 303
```

```
RESULT  72
```



```
HPY239693
LOCUS       HPY239693     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain BO265.
ACCESSION   AJ239693
VERSION     AJ239693.1  GI:4467704
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1..413
                     /organism="Helicobacter pylori"
                     /strain="BO265"
                     /db_xref="taxon:210"
                     /country="Germany"
     gene            1..413
                     /gene="cagA"
     CDS             <1..>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB37804.1"
                     /db_xref="GI:4467705"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVAQFDPDQKPIVDKNDRDNRQ
                     AFDGISQLREEYSNKAIKNPAKKNQYFSDFINKSNDLINKDNLIDVDSSTKSFQKFGD
                     QRYRIFTSWVSHQNDPSKINTRSIQNFMENIIQPP"
BASE COUNT      148 a     82 c     68 g    115 t
ORIGIN
```

```
  Query Match             2.2%;  Score 130;  DB 2;  Length 413;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-50;
  Matches  180;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    769 aagaatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaat 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db    199 AAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAAT 258

Qy    829 ctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccga 888
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    259 CTCATTGATGTAGATTCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGA 318

Qy    889 attttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatc 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 ATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATC 378

Qy    949 c 949
          |
Db    379 C 379
```

```
RESULT  73
HPY239694
LOCUS       HPY239694     413 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain NCTC11637.
ACCESSION   AJ239694
VERSION     AJ239694.1  GI:4467706
KEYWORDS    cagA gene.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 413)
  AUTHORS   Suerbaum,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1999) Suerbaum S., Medical Microbiology,
            Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum,
            GERMANY
FEATURES             Location/Qualifiers
     source          1..413
                     /organism="Helicobacter pylori"
                     /strain="NCTC 11637"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1..413
                     /gene="cagA"
     CDS             <1..>413
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="CagA protein"
                     /protein_id="CAB37805.1"
                     /db_xref="GI:4467707"
                     /translation="EAAFNPQQFINNLQVAFLKVDNAVASYDPDQKPIVDKNDRDNRQ
                     AFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDNLIDIGSSNKSFQKFGT
                     QRYRIFTSWVSHQNDPSKINTRSIRNFMGNIIQPP"
BASE COUNT      148 a     85 c     67 g    113 t
ORIGIN


  Query Match           2.1%;  Score 126;  DB 2;  Length 413;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-48;
  Matches  126;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     712 ggaatcttcgcaatttaagggaagaatactccaatataaagcgatcaaaaatcctaccaaaaaag 771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142 GGAATCTCGCAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAG 201

Qy     772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     202 AATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTC 261

Qy     832 attgat 837
           ||||||
Db     262 ATTGAT 267


RESULT  74
HPY269890
LOCUS       HPY269890     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            AU21 from Australia.
ACCESSION   AJ269890
VERSION     AJ269890.1  GI:5921357
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
```



```
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="AU21"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56452.1"
                     /db_xref="GI:5921358"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDSDQK
                     PIIDKNDRDNRQAFDRISQLREEYSNKAIKNPTKKNQYFSDFIDKSNDLINKDNLIDV
                     ESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      169 a     95 c     68 g    117 t
ORIGIN


  Query Match           2.1%;  Score 125;  DB 2;  Length 449;
  Best Local Similarity  99.4%;  Pred. No. 4.6e-48;
  Matches  175;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     808 gatttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaa 867
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     274 GATTTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAA 333

Qy     868 tttggggatcagcgttaccgaatttttcacaagttgggtgtcccatcaaaacgatccgtct 927
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     334 TTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAATGATCCGTCT 393

Qy     928 aaaatcaacacccgatcgatccgaaattttatggaaaatatcataca 973
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     394 AAAATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATACA 439


RESULT  75
AF198475
LOCUS       AF198475      290 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain Peru34B cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198475
VERSION     AF198475.1  GI:7248676
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
```



```
REFERENCE   1  (bases 1 to 290)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 290)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 290)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .290
                     /organism="Helicobacter pylori"
                     /strain="Peru34B"
                     /db_xref="taxon:210"
                     /country="Peru"
                     /note="European type"
     gene            <1. .>290
                     /gene="cagA"
     CDS             <1. .>290
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43615.1"
                     /db_xref="GI:7248677"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGLKQALSQEEIQNKVDFMEMLAQNNAKLDNLSEK"
BASE COUNT      106 a     51 c     70 g     63 t
ORIGIN


  Query Match           2.1%;  Score 124;  DB 1;  Length 290;
```



```
  Best Local Similarity  100.0%;  Pred. No. 1.4e-47;
  Matches  124;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1561 aacaatgtggctacaatattaattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy    1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 187

Qy    1681 ggct 1684
           ||||
Db     188 GGCT 191


RESULT  76
AF289411
LOCUS       AF289411      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR21 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289411
VERSION     AF289411.1  GI:9964176
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR21"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
```

```
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09860.1"
                     /db_xref="GI:9964177"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTKSFQKFGNQR
                     YQFFTSWVSHQNDPSKINTRSIRNFMEHTIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      111 a     62 c     40 g     78 t
ORIGIN


  Query Match           2.1%;  Score 123;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-47;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    751 atcaaaaatcctaccaaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 60

Qy    811 ttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaattt 870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTT 120

Qy    871 ggg 873
          |||
Db    121 GGG 123


RESULT  77
HPY269885
LOCUS       HPY269885     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            NL105 from The Netherlands.
ACCESSION   AJ269885
VERSION     AJ269885.1  GI:5921347
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
```



```
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="NL105"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56497.1"
                     /db_xref="GI:5921348"
                     /translation="MTNETIDQQPRTEAAFNPQQFINNLQVAFLKVDNVVASFDPNQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPAKKNQYFSDFINKSNDLINKDNLIDV
                     ESSTKSFQKFGDQRYQIPTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      166 a     94 c     70 g    119 t
ORIGIN


  Query Match           2.0%;  Score 121;  DB 2;  Length 449;
  Best Local Similarity   99.4%;  Pred. No. 3.7e-46;
  Matches  171;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    812 taatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaatttg 871
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 TAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTG 337

Qy    872 gggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaaa 931
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    338 GGGATCAGCGTTACCAAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAA 397

Qy    932 tcaacacccgatcgatccgaaatttttatggaaaatatcatacaacccccctat 983
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| .
Db    398 TCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATACAACCCCCTAT 449


RESULT  78
HPY239729
LOCUS       HPY239729     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain NCTC11637.
ACCESSION   AJ239729
```



```
VERSION     AJ239729.1  GI:4467776
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Helicobacter pylori"
                     /strain="NCTC 11637"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1. .243
                     /gene="cagA"
     CDS             <1. .>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37821.1"
                     /db_xref="GI:4467777"
                     /translation="VGYKNQQGDNVATLINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIRNKIDPMEFLAQNNAKLDNLSE"
BASE COUNT       92 a     45 c     51 g     55 t
ORIGIN


  Query Match           2.0%;  Score 119;  DB 2;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-45;
  Matches  119;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1579 attaatgtgcatatgaaaaacggcagtggcttagtcatagcaggtggtgagaaagggatt 1638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     43 ATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCAGGTGGTGAGAAAGGGATT 102

Qy   1639 aacaaccctagtttttatctctacaaagaagaccaactcacaggctcacaacgagcatt 1697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    103 AACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATT 161
```



```
RESULT  79
HPY269897
LOCUS       HPY269897     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            AU10 from Australia.
ACCESSION   AJ269897
VERSION     AJ269897.1  GI:5921371
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="AU10"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56457.1"
                     /db_xref="GI:5921372"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFLKVDNAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDALIDV
                     ESSTKSPQKFWDQRYRIPTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      165 a     98 c     70 g    116 t
ORIGIN


  Query Match           2.0%;  Score 119;  DB 2;  Length 449;
  Best Local Similarity   98.9%;  Pred. No. 3.3e-45;
  Matches  269;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy     554 accaacaaccacaaaaccgaagcggctttaacccgcagcaatttatcaataatcttcaag 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 ACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAG 79

Qy     614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 TGGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCAATCG 139

Qy     674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
           |||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||||
Db     140 TTGATAAGAACGATAGGGATAATAGGCAAGCTTTTGATGGAATCTCGCAATTAAGGGAAG 199

Qy     734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 AATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTA 259

Qy     794 tcaataagagcaatgatttaatcaacaaagac 825
           |||||||||||||||||||||||||||||||
Db     260 TCAATAAGAGCAATGATTTAATCAACAAAGAC 291
```

```
RESULT  80
AF289407
LOCUS       AF289407      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR12 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289407
VERSION     AF289407.1  GI:9964168
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR12"
```



```
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09856.1"
                     /db_xref="GI:9964169"
                     /translation="IENPTKKNQYPSDPINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDREKAEFLKSAKQSPAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      107 a     60 c     46 g     78 t
ORIGIN
```

```
  Query Match             2.0%;  Score 118;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 1e-44;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     868 tttggggatcagcgttaccgaatttttcacaagttgggtgtcccatcaaaacgatccgtct 927
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 TTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCT 177

Qy     928 aaaatcaacacccgatcgatccgaaatttttatggaaaaatatcatacaaccccctatcc 985
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 AAAATCAACACCCGATCGATCCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 235
```

```
RESULT  81
AF289456
LOCUS       AF289456      678 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR52 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289456
VERSION     AF289456.1  GI:9964266
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
```



```
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Helicobacter pylori"
                     /isolate="CR52"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>678
                     /gene="cagA"
     CDS             <1. .>678
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09905.1"
                     /db_xref="GI:9964267"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKAGQA
                     ASPEEPIYAQVAKKVKAKIDQLNQAASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDD
                     LSKVGLSRNQELAQKIDNLNQAVSEAKAGFFGNLEQTIDKLKDSIKHNPMNLWAESAK
                     KMPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      236 a    126 c    150 g    166 t
ORIGIN
```

```
  Query Match             2.0%;  Score 118;  DB 2;  Length 678;
  Best Local Similarity   100.0%;  Pred. No. 9.6e-45;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3406 ctttcaaggaatcaagaattggctcagaaaaattgacaatctcaatcaagcggtatcagaa 3465
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     496 CTTTCAAGGAATCAAGAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAA 555

Qy    3466 gctaaagcaggttttttttggcaatctagagcaaacgatagacaagctcaaagattcta 3523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     556 GCTAAAGCAGGTTTTTTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTA 613
```



```
RESULT  82
AF289437
LOCUS       AF289437      540 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR7 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289437
VERSION     AF289437.1  GI:9964228
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 540)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Helicobacter pylori"
                     /isolate="CR7"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>540
                     /gene="cagA"
     CDS             <1. .>540
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09886.1"
                     /db_xref="GI:9964229"
                     /translation="NFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKAGQAASP
                     EEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKGHDKVDDLSKVGRSVSPEPI
                     YATIDDLGGSSTLNKVGLSRNQKLVQKIDNLNQAVSEAKAGFFGNLEQMIDKLKDSTK
                     HNPMNLWAEGAKKVPASLSA"
BASE COUNT      195 a     98 c    118 g    129 t
ORIGIN
```

```
  Query Match             2.0%;  Score 116;  DB 2;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-44;
  Matches  116;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3124 aacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaataac 3183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AACTTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAATAAC 60

Qy   3184 aataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcagg 3239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGG 116


RESULT  83
AF289450
LOCUS       AF289450      567 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR42 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289450
VERSION     AF289450.1  GI:9964254
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Helicobacter pylori"
                     /isolate="CR42"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>567
                     /gene="cagA"
     CDS             <1. .>567
```



```
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09899.1"
                     /db_xref="GI:9964255"
                     /translation="NFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKAGQVASP
                     EEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSPEPI
                     YATIDDLGGPFPLKRHDKVDDLSKVGLSREQQLKQKIDNLSQAVSEAKAGFFGNLEQT
                     IDKLKDSTKNNTVNLWAESAKKVPASLSA"
BASE COUNT      209 a     98 c    124 g    136 t
ORIGIN


  Query Match             2.0%;  Score 116;  DB 2;  Length 567;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-44;
  Matches  116;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3124 aacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaataac 3183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AACTTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAATAAC 60

Qy   3184 aataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcagg 3239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGG 116


RESULT  84
AF289457
LOCUS       AF289457      567 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR53 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289457
VERSION     AF289457.1  GI:9964268
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Helicobacter pylori"
                     /isolate="CR53"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>567
                     /gene="cagA"
     CDS             <1. .>567
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09906.1"
                     /db_xref="GI:9964269"
                     /translation="NFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKAGQVASP
                     EEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSPEPI
                     YATIDDLGGPFPLKRHDKVDDLSKVGLSRNQELTQKIDNLSQAVSEAKAGYFGNLEQT
                     IDNLKDSTKNNPMNLWVESAKKVPACLSA"
BASE COUNT      208 a    100 c    121 g    138 t
ORIGIN


  Query Match             2.0%;  Score 116;  DB 2;  Length 567;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-44;
  Matches  116;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3124 aacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaataac 3183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AACTTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAATAAC 60

Qy   3184 aataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcagg 3239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGG 116


RESULT  85
AF198476
LOCUS       AF198476      284 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain Peru35B cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198476
VERSION     AF198476.1  GI:7248678
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
```



```
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 284)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 284)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 284)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .284
                     /organism="Helicobacter pylori"
                     /strain="Peru35B"
                     /db_xref="taxon:210"
                     /country="Peru"
                     /note="European type"
     gene            <1. .>284
                     /gene="cagA"
     CDS             <1. .>284
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43616.1"
                     /db_xref="GI:7248679"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGSQRALSQEEIQNKIDFMEFLAQNNAKLDNLS"
BASE COUNT      103 a     53 c     64 g     64 t
ORIGIN
```

```
  Query Match             1.9%;  Score 115;  DB 1;  Length 284;
  Best Local Similarity   99.1%;  Pred. No. 2.7e-43;
  Matches  215;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    128 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTACCTCTACAAAGAAGACCAACTCACA 187

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    188 GGCTCACAACGAGCATTGAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 247

Qy   1741 cttgcacaaaataatgctaaattagacaacttgagcg 1777
          |||||||||||||||||||||||||||||||||||||
Db    248 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCG 284


RESULT  86
AF190660
LOCUS       AF190660     2963 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain 84-183 cag pathogenicity island
            polymorphic right end, type Ia motif; insertion sequence IS606,
            complete sequence; and glutamate racemase (glr) gene, partial cds.
ACCESSION   AF190660
VERSION     AF190660.1  GI:10567547
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 2963)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 2963)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (29-SEP-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .2963
                     /organism="Helicobacter pylori"
                     /strain="84-183"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="USA"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ic motif; similar to
                     Helicobacter pylori strain Gambia94/24 type Ia motif
                     presented in GenBank Accession Number AF190658"
     misc_feature    26. .157
                     /note="similar to Helicobacter pylori insertion sequence
                     IS606"
     gene            210. .640
                     /gene="putative DNA helicase"
                     /note="truncated; similar to Helicobacter pylori strain
                     26695 putative helicase gene HP0548"
     source          739. .2705
                     /organism="Helicobacter pylori"
                     /strain="84-183"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /insertion_seq="IS606"
                     /country="USA"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ic motif; similar to type Ia
                     motif; similar to Helicobacter pylori insertion sequence
                     IS606 presented in GenBank Accession Number U95957"
     repeat_region   2765. .2863
                     /note="defines the terminus of the cag pathogenicity
                     island"
                     /rpt_type=direct
     gene            complement(2765. .>2963)
                     /gene="glr"
                     /note="flanks cag pathogenicity island"
     CDS             complement(2765. .>2963)
                     /gene="glr"
                     /note="similar to Helicobacter pylori strain 26695
                     glutamate racemase HP0549"
                     /codon_start=2
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG18483.1"
                     /db_xref="GI:10567548"
                     /translation="IEGYFMGHFALPTPPLLIHSGDAIVEYLQQKYALKKNAHAFPKV
                     EFHASGDVIWLEKQAKEWLKL"
BASE COUNT     1027 a    566 c    521 g    849 t
```



```
ORIGIN


  Query Match             1.9%;  Score 114;  DB 1;  Length 2963;
  Best Local Similarity   99.4%;  Pred. No. 7e-43;
  Matches  164;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   4474 aactagaaaaaaacaagtagctataacaaagatcaagttcaaaaaatcatagagctttttag 4533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    222 AACTAGAAAAAACAAGTAGCTATAACAAAGATCAAGTTCAAAAAATCATAGAGCTTTTAG 281

Qy   4534 agcaaattgatcgcgctcttaaccaaagaaaaatcagaaaaaccatagaattatcacac 4593
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    282 AGCAAATCGATCGCGCTCTTAACCAAAGAAAAATCAGAAAAACCATAGGAATTATCACAC 341

Qy   4594 cttataatgcccaaaaaagacgcttgcgatcagaagtggaaaaat 4638
          |||||||||||||||||||||||||||||||||||||||||||||
Db    342 CTTATAATGCCCAAAAAAGACGCTTGCGATCAGAAGTGGAAAAAT 386


RESULT  87
AF289399
LOCUS       AF289399      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR1 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289399
VERSION     AF289399.1  GI:9964152
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR1"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
```



```
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09848.1"
                     /db_xref="GI:9964153"
                     /translation="IKNPTKKNQYFSDFINKINDLINKDNLIDVESSTKSFQKFGDQR
                     YQIFTSWVSHQNDPSKINTRSIRNFMEHTIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      110 a     62 c     41 g     78 t
ORIGIN


  Query Match             1.9%;  Score 111;  DB 2;  Length 291;
  Best Local Similarity   99.1%;  Pred. No. 2.2e-41;
  Matches  211;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGATCAATGAT 60

Qy    811 ttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaattt 870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTT 120

Qy    871 ggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaa 930
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGGATCAGCGTTACCAAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAA 180

Qy    931 atcaacacccgatcgatccgaaattttatggaa 963
          |||||||||||||||||||||||||||||||||
Db    181 ATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT  88
AF289416
LOCUS       AF289416      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR38 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289416
VERSION     AF289416.1  GI:9964186
KEYWORDS    .
SOURCE      Helicobacter pylori.
```

```
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR38"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09865.1"
                     /db_xref="GI:9964187"
                     /translation="IKNPTKKNQYFSDFINKINDLINKDNLIDVESSTKSPQKFGDQR
                     YQIFTSWVSHQNDPSKINTRSIRNFMEHTIQPPIPDDKEKAEPLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      110 a      62 c      41 g      78 t
ORIGIN


  Query Match             1.9%;  Score 111;  DB 2;  Length 291;
  Best Local Similarity   99.1%;  Pred. No. 2.2e-41;
  Matches  211;  Conservative     0;  Mismatches     2;  Indels     0;  Gaps     0;

Qy     751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGATCAATGAT 60
```



```
Qy     811 ttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaattt 870
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTT 120

Qy     871 ggggatcagcgttaccgaatttttcacaagttgggtgtcccatcaaaacgatccgtctaaa 930
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGGGATCAGCGTTACCAAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAA 180

Qy     931 atcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||
Db     181 ATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT  89
AF289418
LOCUS       AF289418      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR43 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289418
VERSION     AF289418.1  GI:9964190
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR43"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
```



```
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09867.1"
                     /db_xref="GI:9964191"
                     /translation="VKNPTKKNQYFSDFINKSNDLINKDALIDVESSTESPQKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHIIQPPIPDDKEKAEPLKSAKQSPAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      104 a      64 c      47 g      76 t
ORIGIN


  Query Match             1.9%;  Score 110;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-41;
  Matches  110;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy     854 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104 AGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATC 163

Qy     914 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 AAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT  90
AF289429
LOCUS       AF289429      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR58 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289429
VERSION     AF289429.1  GI:9964212
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
```



```
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR58".
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09878.1"
                     /db_xref="GI:9964213"
                     /translation="IKNPTKKNQYFSDFINKINDLINKDNLIDVESSTKSFQKFGDQR
                     YQIFTSWVSHQNDPSKINTRSIRNFMEHIIQPPIHDDKEKAEFLKSAKQSPAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      111 a      62 c      41 g      77 t
ORIGIN


  Query Match             1.8%;  Score 109;  DB 2;  Length 291;
  Best Local Similarity   99.4%;  Pred. No. 2e-40;
  Matches  159;  Conservative     0;  Mismatches     1;  Indels     0;  Gaps     0;

Qy     804 caatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttca 863
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 CAATGATTTAATCAACAAAGACAATCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCA 113

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCAAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           ||||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT  91
AF190658
LOCUS       AF190658      987 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain Gambia94/24 cag pathogenicity island
            polymorphic right end, type Ia motif, and glutamate racemase (glr)
            gene, partial cds.
```

```
ACCESSION   AF190658
VERSION     AF190658.1  GI:10567543
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 987)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="Helicobacter pylori"
                     /strain="Gambia94/24"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ia motif"
     misc_feature    28. .160
                     /note="similar to Helicobacter pylori insertion sequence
                     IS606 presented in GenBank Accession Number U95957"
     gene            216. .641
                     /gene="putative DNA helicase"
                     /note="truncated; similar to Helicobacter pylori strain
                     26695 putative helicase gene HP0548"
     gene            complement(804. .>987)
                     /gene="glr"
                     /note="flanks cag pathogenicity island"
     repeat_region   804. .987
                     /note="defines the terminus of the cag pathogenicity
                     island"
                     /rpt_type=direct
     CDS             complement(804. .>987)
                     /gene="glr"
                     /note="similar to Helicobacter pylori strain 26695
```



```
                     glutamate racemase HP0549"
                     /codon_start=2
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG18481.1"
                     /db_xref="GI:10567544"
                     /translation="MGHFALPTPPLLIHSGDAIVEYLQQKYALKKNAHAFPKVEPHAS
                     GDVIWLEKQAKEWLKL"
BASE COUNT      330 a    181 c    183 g    293 t
ORIGIN


  Query Match             1.8%;  Score 109;  DB 1;  Length 987;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-40;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5034 taaatttttattagcgttacaatttgagccattctttagcttgtttttctagccagatca 5093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     788 TAAATTTTTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCA 847

Qy    5094 catcgccgctcgcatgaaattccacttttagggaatgcgtgtgcattttt 5142
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     848 CATCGCCGCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATTTTT 896


RESULT  92
HPY269875
LOCUS       HPY269875     464 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            US238 from United States.
ACCESSION   AJ269875
VERSION     AJ269875.1  GI:5921327
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 464)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .464
```



```
                     /organism="Helicobacter pylori"
                     /isolate="US218"
                     /db_xref="taxon:210"
                     /country="USA"
     gene            1. .464
                     /gene="cagA"
     CDS             1. .>464
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56487.1"
                     /db_xref="GI:5921328"
                     /translation="MTNETIDQTRTPDQTSSQTAFDPQRFINNLQVAFLKVDSAVASF
                     DPDQKPIVDKNDRDNRQAFEKISQLREEFANKAIKNPTKKNQYFSNFISKSNDLINKD
                     GLIDTGSSIKSFQKFGDQRYRIFTSWVSHQNDPSKINTRSIRNFMERTIQPP"
BASE COUNT      171 a     97 c     75 g    121 t
ORIGIN


  Query Match             1.8%;  Score 108;  DB 2;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-40;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     856 agctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaa 915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     337 AGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAA 396

Qy     916 aacgatccgtctaaaatcaacacccgatcgatccgaaatttttatggaa 963
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db     397 AACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 444


RESULT  93
AF222807
LOCUS       AF222807     1335 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain India41 CagA (cagA) gene, partial cds.
ACCESSION   AF222807
VERSION     AF222807.1  GI:7021426
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 1335)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
            Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
            Nair,G.B. and Berg,D.E.
  TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
            India
  JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
```



```
   MEDLINE   20270153
    PUBMED   10809703
REFERENCE   2  (bases 1 to 1335)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Datta,S., Chowdhury,A., Nair,G.B.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Molecular Microbiology, Washington
            University Medical School, 660 South Euclid Avenue, Saint Louis, MO
            63108, USA
FEATURES             Location/Qualifiers
     source          1. .1335
                     /organism="Helicobacter pylori"
                     /strain="India41"
                     /db_xref="taxon:210"
     gene            <1. .>1335
                     /gene="cagA"
     CDS             <1. .>1335
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA"
                     /protein_id="AAF35340.1"
                     /db_xref="GI:7021427"
                     /translation="IWLENSVKDVIINQKITDKVDNLNQAVSVAKATGDFSRVEQALA
                     DLNNFSKEQLAQQAQKNESFNAGKKSEIYQSVKNGVNGTLVGNGLSGIEATALAKNFS
                     DIKKELNEKFKNFNNNNNNGLKNSTEPIYAKVNKKTQGQAASLEEPIYAQVAKKVNA
                     KIDRLNQIASGLGVVGQAAGPPLKKHDKVDDLSKVGLSASPEPIYASIDDLGGPFPLK
                     RHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGLSRNQKLAQKID
                     NLNQAVSEAKAGFFGNLEQTIDNLKDSAKDNPVNLWVESAKKVPASLSAKLDNYATNG
                     HTRINSNIKNGAINEKATGMLTQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKN
                     MKDYSDSFKFSTKLNNAAKDTKSGFTQFLTNAFSTASYYCLAEEHAEHGIKNA"
BASE COUNT      476 a    251 c    278 g    330 t
ORIGIN


  Query Match             1.8%;  Score 108;  DB 1;  Length 1335;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-40;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3621 acgcattaatagcaatatcaaaaatggagcaatcaatgaaaaagcgaccggcatgctaac 3680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1008 ACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAAAAAGCGACCGGCATGCTAAC 1067

Qy    3681 gcaaaaaaaccctgagtggctcaagctcgtgaatgataagatagttgc 3728
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1068 GCAAAAAAACCCTGAGTGGCTCAAGCTCGTGAATGATAAGATAGTTGC 1115


RESULT  94
AF289415
LOCUS       AF289415     291 bp    DNA             BCT       02-SEP-2000
```

```
DEFINITION  Helicobacter pylori isolate CR35 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289415
VERSION     AF289415.1  GI:9964184
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR35"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09864.1"
                     /db_xref="GI:9964185"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDALIDVESSTKSFQKFGDQR
                     YRIFASWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      107 a      63 c      45 g      76 t
ORIGIN

  Query Match             1.8%;  Score 107;  DB 2;  Length 291;
  Best Local Similarity   99.4%;  Pred. No. 1.7e-39;
```



```
  Matches  157;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     828 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 TCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCG 137

Qy     888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 AATTTTCGCAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 197

Qy     948 ccgaaatttttatggaaaaatatcatacaaccccctatcc 985
           |||||||||||||||||||||||||||||||||||||||
Db     198 CCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 235


RESULT  95
AF289430
LOCUS       AF289430      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR59 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289430
VERSION     AF289430.1  GI:9964214
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR59"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
```



```
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09879.1"
                     /db_xref="GI:9964215"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDALIDVESSTKSFQKFGDQR
                     YRIFTSWVSHQNDPSKIDTRSIRNFMENIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      107 a      63 c      45 g      76 t
ORIGIN

  Query Match             1.8%;  Score 107;  DB 2;  Length 291;
  Best Local Similarity   99.4%;  Pred. No. 1.7e-39;
  Matches  157;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     828 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 TCTCATTGATGTAGAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCG 137

Qy     888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     138 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 197

Qy     948 ccgaaatttttatggaaaatatcatacaaccccctatcc 985
           |||||||||||||||||||||||||||||||||||||||
Db     198 CCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 235


RESULT  96
AF190996
LOCUS       AF190996      590 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain OhioP46 cag pathogenicity island
            polymorphic right end, type Ic motif right-end.
ACCESSION   AF190996
VERSION     AF190996.1  GI:10567559
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
```



```
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
  PUBMED    10809702
REFERENCE   2  (bases 1 to 590)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Helicobacter pylori"
                     /strain="OhioP46"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="USA: Ohio"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ic motif right-end; similar to
                     Helicobacter pylori strain 84-183 type Ic motif presented
                     in GenBank Accession Number AF190660; truncated inside of
                     full copy of IS606"
BASE COUNT      175 a      102 c      127 g      186 t
ORIGIN

  Query Match             1.8%;  Score 107;  DB 1;  Length 590;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-39;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5033 ttaaatttttattagcgttacaatttgagccattctttagcttgtttttctagccagatc 5092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     367 TTAAATTTTTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATC 426

Qy    5093 acatcgccgctcgcatgaaattccactttagggaatgcgtgtgcatt 5139
           |||||||||||||||||||||||||||||||||||||||||||||
Db     427 ACATCGCCGCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATT 473


RESULT  97
AF190992
LOCUS       AF190992      956 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain OhioM6 cag pathogenicity island
            polymorphic right end, type Ia motif.
ACCESSION   AF190992
VERSION     AF190992.1  GI:10567555
KEYWORDS    .
SOURCE      Helicobacter pylori.
```

```
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 956)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 956)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="Helicobacter pylori"
                     /strain="OhioM6"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="USA: Ohio"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island: type Ia motif; similar to
                     Helicobacter pylori strain Gambia94/24 type Ia motif
                     presented in GenBank Accession Number AF190658"
BASE COUNT      316 a    177 c    176 g    287 t
ORIGIN


  Query Match             1.8%;  Score 104;  DB 1;  Length 956;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-38;
  Matches  104;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5039 ttttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcg 5098
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    763 TTTTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCG 822

Qy   5099 ccgctcgcatgaaatttccactttagggaatgcgtgtgcattttt 5142
          |||||||||||||||||||||||||||||||||||||||||||
Db    823 CCGCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATTTTT 866


RESULT  98
```



```
AF289442
LOCUS       AF289442      567 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR19 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289442
VERSION     AF289442.1  GI:9964238
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Helicobacter pylori"
                     /isolate="CR19"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>567
                     /gene="cagA"
     CDS             <1. .>567
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09891.1"
                     /db_xref="GI:9964239"
                     /translation="NFSDIKKELNAKLRNFNNNNNNGLKNEPIYAKVNKKKTGQVASP
                     EEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSPEPI
                     YATIDDLGGPFPLKRHDKVDDLSKVGLSREQELAQKIDNLNQAVSEAKAGYFGNLERT
                     IDKLKDSTKHNPMNLWVESAKKVPANLSA"
BASE COUNT      207 a    102 c    124 g    134 t
ORIGIN


  Query Match             1.7%;  Score 101;  DB 2;  Length 567;
```



```
  Best Local Similarity   99.3%;  Pred. No. 1.2e-36;
  Matches  151;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3255 tgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaa 3314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 TGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAA 191

Qy   3315 tcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttccctttgaaaaggca 3374
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    192 TCAAATAGCAAGTGGTTTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCA 251

Qy   3375 tgataaagttgatgatctcagtaaggtagggc 3406
          |||||||||||||||||||||||||||||||
Db    252 TGATAAAGTTGATGATCTCAGTAAGGTAGGGC 283


RESULT  99
AF289435
LOCUS       AF289435      678 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR5 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289435
VERSION     AF289435.1  GI:9964224
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Helicobacter pylori"
                     /isolate="CR5"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>678
                     /gene="cagA"
```



```
     CDS             <1. .>678
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09884.1"
                     /db_xref="GI:9964225"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDD
                     LSKVGRSREQQLKQKIDNLSQAVSEAKAGPFGNLEQTIDKLKDSTKNNTVNLWAESAK
                     KVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      239 a    122 c    152 g    165 t
ORIGIN


  Query Match             1.7%;  Score 101;  DB 2;  Length 678;
  Best Local Similarity   99.3%;  Pred. No. 1.2e-36;
  Matches  151;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3255 tgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaa 3314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAA 200

Qy   3315 tcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttccctttgaaaaggca 3374
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    201 TCAAATAGCAAGTGGTTTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCA 260

Qy   3375 tgataaagttgatgatctcagtaaggtagggc 3406
          |||||||||||||||||||||||||||||||
Db    261 TGATAAAGTTGATGATCTCAGTAAGGTAGGGC 292


RESULT 100
AF222808
LOCUS       AF222808     1224 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain India91 CagA (cagA) gene, partial cds.
ACCESSION   AF222808
VERSION     AF222808.1  GI:7021428
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 1224)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
            Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
```

```
            Nair,G.B. and Berg,D.E.
  TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
            India
  JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
  MEDLINE   20270153
   PUBMED   10809703
REFERENCE   2  (bases 1 to 1224)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Datta,S., Chowdhury,A., Nair,G.B.
            and Berg,D.E.
  TITLE     Direct Submiasion
  JOURNAL   Submitted (10-JAN-2000) Molecular Microbiology, Washington
            University Medical School, 660 South Euclid Avenue, Saint Louis, MO
            63108, USA
FEATURES             Location/Qualifiers
     source          1. .1224
                     /organism="Helicobacter pylori"
                     /strain="India91"
                     /db_xref="taxon:210"
     gene            <1. .>1224
                     /gene="cagA"
     CDS             <1. .>1224
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA"
                     /protein_id="AAF35341.1"
                     /db_xref="GI:7021429"
                     /translation="SDLENSVKDVIINQKVTDKVDNLSSAVSVAKATGDFSRVEQALA
                     DLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNSVNGTLVGNGLSQAEATTLSKNFS
                     DIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKTQGQAASLEEPIYAQVAKKVNAKI
                     DRLNQIASGLGGVGQAAGFPLLKKHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRH
                     DKVDDLSKVGLSREQQLAQKIDNLNQAVSEAKAGFPGNLEQTINKLKDSTKHNPMNLW
                     VESAKQVPASLSAKLDNYATNSHTRINSNIKSGAINEKATGMLTQKNPEWLKLVNDKI
                     VAHNVGSVPLSEYDKIGFNQRNMKDYSDSFKFSTKLNNAAKDIKSGFTQFLANAFSTG
                     YTCLAGENAEHGIKNV"
BASE COUNT      438 a    227 c    262 g    297 t
ORIGIN
```

```
  Query Match             1.7%;  Score 101;  DB 1;  Length 1224;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-36;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3237 agggcaagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgc 3296
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    414 AGGGCAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGC 473

Qy   3297 aaaaattgaccgactcaatcaaatagcaagtggtttgggtg 3337
          |||||||||||||||||||||||||||||||||||||||||
Db    474 AAAAATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTG 514
```



```
RESULT 101
AF289403
LOCUS       AF289403      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR6 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289403
VERSION     AF289403.1  GI:9964160
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR6"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09852.1"
                     /db_xref="GI:9964161"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESPRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a     62 c     46 g     77 t
ORIGIN
```



```
  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy    924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
          ||||||||||||||||||||||||||||||||||||||||
Db    174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 102
AF289405
LOCUS       AF289405      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR8 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289405
VERSION     AF289405.1  GI:9964164
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR8"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
```



```
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09854.1"
                     /db_xref="GI:9964165"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSVESPRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHTIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      105 a     61 c     47 g     78 t
ORIGIN
```

```
  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy    924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
          ||||||||||||||||||||||||||||||||||||||||
Db    174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 103
AF289409
LOCUS       AF289409      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR19 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289409
VERSION     AF289409.1  GI:9964172
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR19"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09858.1"
                     /db_xref="GI:9964173"
                     /translation="IKNPTKKNQYFSDFIEKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHTIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      107 a      63 c      46 g      75 t
ORIGIN


  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 104
AF289410
LOCUS       AF289410      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR20 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289410
VERSION     AF289410.1  GI:9964174
KEYWORDS    .
```



```
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR20"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09859.1"
                     /db_xref="GI:9964175"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a      62 c      46 g      77 t
ORIGIN


  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 105
AF289413
LOCUS       AF289413      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR23 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289413
VERSION     AF289413.1  GI:9964180
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR23"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09862.1"
                     /db_xref="GI:9964181"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHTIQPPIPDDKEKAEFLKSAKQSFAG"
```



```
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a      62 c      46 g      77 t
ORIGIN


  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 106
AF289419
LOCUS       AF289419      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR48 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289419
VERSION     AF289419.1  GI:9964192
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR48"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
```

```
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09868.1"
                     /db_xref="GI:9964193"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEYIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a      61 c      46 g      78 t
ORIGIN


  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 107
AF289421
LOCUS       AF289421      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR50 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289421
VERSION     AF289421.1  GI:9964196
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
```



```
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR50"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09870.1"
                     /db_xref="GI:9964197"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a      62 c      46 g      77 t
ORIGIN


  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 108
AF289425
LOCUS       AF289425      291 bp    DNA             BCT       02-SEP-2000
```



```
DEFINITION  Helicobacter pylori isolate CR54 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289425
VERSION     AF289425.1  GI:9964204
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR54"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09874.1"
                     /db_xref="GI:9964205"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a      62 c      46 g      77 t
ORIGIN


  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
```



```
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 109
AF289427
LOCUS       AF289427      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR56 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289427
VERSION     AF289427.1  GI:9964208
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR56"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09876.1"
                     /db_xref="GI:9964209"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHIRQPPIPDDKEKAEFLKSAKQSFAG"
       misc_feature  115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a      62 c      47 g      76 t
ORIGIN

  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 110
AF289428
LOCUS       AF289428      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR57 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289428
VERSION     AF289428.1  GI:9964210
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
```



```
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR57"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09877.1"
                     /db_xref="GI:9964211"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRNFMEHIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a      62 c      46 g      77 t
ORIGIN

  Query Match             1.7%;  Score 100;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-36;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy     924 gtctaaaatcaacacccgatcgatccgaaattttatggaa 963
           |||||||||||||||||||||||||||||||||||||||
Db     174 GTCTAAAATCAACACCCGATCGATCCGAAATTTTATGGAA 213


RESULT 111
AF289408
LOCUS       AF289408      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR13 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289408
VERSION     AF289408.1  GI:9964170
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
```



```
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR13"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09857.1"
                     /db_xref="GI:9964171"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVKSSTESFRKFGDQR
                     YQIFTSWVSHQNDPSKINTRSIRNFMENIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      109 a      60 c      44 g      78 t
ORIGIN

  Query Match             1.7%;  Score 98;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-35;
  Matches   98;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 197

Qy     948 ccgaaatttttatggaaaatatcatacaaccccctatcc 985
           |||||||||||||||||||||||||||||||||||||||
Db     198 CCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 235
```



```
RESULT 112
AF190995
LOCUS       AF190995      897 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain OhioP46 cag pathogenicity island
            polymorphic right end, type Ic motif left-end.
ACCESSION   AF190995
VERSION     AF190995.1  GI:10567558
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 897)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Helicobacter pylori"
                     /strain="OhioP46"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="USA: Ohio"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ic motif left-end; similar to
                     Helicobacter pylori strain 84-183 type Ic motif presented
                     in GenBank Accession Number AF190660; truncated inside of
                     full copy of IS606"
BASE COUNT      310 a     170 c     152 g     265 t
ORIGIN

  Query Match             1.6%;  Score 97;  DB 1;  Length 897;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-35;
```

```
  Matches   97;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4542 gatcgcgctcttaaccaaagaaaaatcagaaaaaccataggaattatcacaccttataat 4601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    292 GATCGCGCTCTTAACCAAAGAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAAT 351

Qy   4602 gcccaaaaaagacgcttgcgatcagaagtggaaaaat 4638
          |||||||||||||||||||||||||||||||||||||
Db    352 GCCCAAAAAAGACGCTTGCGATCAGAAGTGGAAAAAT 388


RESULT 113
HPY269892
LOCUS       HPY269892     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate CH7
            from China.
ACCESSION   AJ269892
VERSION     AJ269892.1  GI:5921361
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="CH7"
                     /db_xref="taxon:210"
                     /country="China"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56454.1"
                     /db_xref="GI:5921362"
                     /translation="MTNEPITQQPRTEVAFNPQQFINNLQVAFLKIDNAVTSYDPDQK
```



```
                     PIVNKNDRDNRQAFDGISQLREEYSSKAIRNPTKKNQYFSDFINKSNDLINKDALIEI
                     ESSTKSFQKFGDQRYQIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      165 a    100 c     70 g    114 t
ORIGIN


  Query Match             1.6%;  Score 96;  DB 2;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-34;
  Matches   96;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    354 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 413

Qy    948 ccgaaattttatggaaaatatcatacaaccccctat 983
          |||||||||||||||||||||||||||||||||||||
Db    414 CCGAAATTTTATGGAAAATATCATACAACCCCCTAT 449


RESULT 114
HPY269893
LOCUS       HPY269893     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate CH8
            from China.
ACCESSION   AJ269893
VERSION     AJ269893.1  GI:5921363
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="CH8"
                     /db_xref="taxon:210"
                     /country="China"
     gene            1. .449
                     /gene="cagA"
```



```
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56455.1"
                     /db_xref="GI:5921364"
                     /translation="MTNETINQQPRTEVAFNPQQFINNLQVAFLKIDNAVTSYDPDQK
                     PIVNKNDRDNRQAFDGISQLREEYSSKAIKNPTKKNQYFSDFINRSNDLINKDALIDI
                     ESSTKSFQKFGDQRYQIFTSWVSHQNDPSKINTRSIRNFMENIIQPP"
BASE COUNT      169 a    97 c     68 g    115 t
ORIGIN


  Query Match             1.6%;  Score 96;  DB 2;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-34;
  Matches   96;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    354 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 413

Qy    948 ccgaaattttatggaaaatatcatacaaccccctat 983
          |||||||||||||||||||||||||||||||||||||
Db    414 CCGAAATTTTATGGAAAATATCATACAACCCCCTAT 449


RESULT 115
AF190993
LOCUS       AF190993     1000 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain HUP B3 cag pathogenicity island
            polymorphic right end, type Ia motif.
ACCESSION   AF190993
VERSION     AF190993.1  GI:10567556
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 1000)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
```



```
REFERENCE   2  (bases 1 to 1000)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .1000
                     /organism="Helicobacter pylori"
                     /strain="HUP B3"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /note="isolated from Spanish gastric ulcer patient;
                     polymorphic region at right end of the cag pathogenicity
                     island; type Ia motif; similar to Helicobacter pylori
                     strain Gambia94/24 type Ia motif presented in GenBank
                     Accession Number AF190658"
BASE COUNT      335 a    183 c    185 g    297 t
ORIGIN


  Query Match             1.6%;  Score 96;  DB 1;  Length 1000;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-34;
  Matches   96;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4543 atcgcgctcttaaccaaagaaaaatcagaaaaaccataggaattatcacaccttataatg 4602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 ATCGCGCTCTTAACCAAAGAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATG 359

Qy   4603 cccaaaaaagacgcttgcgatcagaagtggaaaaat 4638
          |||||||||||||||||||||||||||||||||||||
Db    360 CCCAAAAAAGACGCTTGCGATCAGAAGTGGAAAAAT 395


RESULT 116
AF190994
LOCUS       AF190994      940 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain Peru002B cag pathogenicity island
            polymorphic right end, type Ia motif.
ACCESSION   AF190994
VERSION     AF190994.1  GI:10567557
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 940)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
```

```
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 940)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1..940
                     /organism="Helicobacter pylori"
                     /strain="Peru002B"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="Peru"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ia motif; similar to
                     Helicobacter pylori strain Gambia94/24 type Ia motif
                     presented in GenBank Accession Number AF190650"
BASE COUNT      319 a    172 c    177 g    272 t
ORIGIN


  Query Match             1.6%;  Score 94;  DB 1;  Length 940;
  Best Local Similarity   99.3%;  Pred. No. 2.5e-33;
  Matches  144;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   4549 ctcttaaccaaagaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaa 4608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 CTCTTAACCAAAGAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAA 376

Qy   4609 aaagacgcttgcgatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaa 4668
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    377 AAAGACGCTTGCGATCAGAAGTGGAAAAATGCGGCTTCAAGAATTTTGATGAGCTCAAAA 436

Qy   4669 tagacactgtggatgcctttcaagg 4693
           |||||||||||||||||||||||||
Db    437 TAGACACTGTGGATGCCTTTCAAGG 461


RESULT 117
AF202972
LOCUS       AF202972     3522 bp    DNA             BCT       02-MAR-2000
DEFINITION  Helicobacter pylori strain GC401(F32) CagA (cagA) gene, complete
```



```
            cds.
ACCESSION   AF202972
VERSION     AF202972.1  GI:6573219
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 3522)
  AUTHORS   Asahi,M., Azuma,T., Ito,S., Ito,Y., Suto,H., Nagai,Y.,
            Tsubokawa,M., Tohyama,Y., Maeda,S., Omata,M., Suzuki,T. and
            Sasakawa,C.
  TITLE     Helicobacter pylori CagA protein can be tyrosine phosphorylated in
            gastric epithelial cells
  JOURNAL   J. Exp. Med. 191 (4), 593-602 (2000)
  MEDLINE   20150113
REFERENCE   2  (bases 1 to 3522)
  AUTHORS   Ito,Y. and Azuma,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-1999) Second Department of Internal Medicine,
            Fukui Medical University, Matsuoka-cho, Yoshida-gun, Fukui
            910-1193, Japan
FEATURES             Location/Qualifiers
     source          1..3522
                     /organism="Helicobacter pylori"
                     /strain="GC401(F32)"
                     /db_xref="taxon:210"
     gene            1..3522
                     /gene="cagA"
     CDS             1..3522
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA"
                     /protein_id="AAF17597.1"
                     /db_xref="GI:6573220"
                     /translation="MTNETIDQTTTPDQTGFVPQRFINNLQVAPIKVDNAVASFDPDQ
                     KPIVDKNDKDNRQAYEKISQLREEYANKAIKNPAKKNQYFSDFINKSNDLINKDNLIA
                     VDSSVESFRKFGDQRYQIFTSWVSLQKDPSKINTQQIRNFMENVIKPPISDDKEKAEF
                     LRSAKQSFAGIIIGNQIRSDEKFMGVFDESLKARQEAEKNAEPAGGDWLDIFLSFVFN
                     KKQSSDLKETLNQEPRPDFEQNLATTTTDIQGLPPEARDLLDERGNFFKFTLGDVEML
                     DVEGVADKDPNYKFNQLLIHNNALSSMLMGSHSNIEPEKVSLLYGDNGGPEARHDWNA
                     TVGYKNQQGNNVATLINAHLNNGSGLIIAGNEDGIKNPSFFLYKEDQLTGLKQALSQE
                     EIQNKVDFNEPLAQNNAKLDNLSEKEEKFQTEIENPQKDRKAYLDALGNDHIAFVSK
                     KDPKHLALVTEFGNGELSYTLKDYGKKQDKALDGETKTTLQGSLKYDGVMFVNYSNFK
                     YTNASKSPNKGLGTTNGVSHLEANFSKVAVFNLPNLNNLAITNYIRRDLEDKLWAKGL
                     SPQEANKLIKDFLNSNKEMVGKVSNFNKAVAEAKNTGNYDEVKKAQKDLEKSLRRREH
                     LEKEVAKKLESRNDNKNRMEAKAQANSQKDKIFALISQEASKEARVATFDPYLKGVRS
                     ELSDKLENINKNLKDFGKSFDELKSGKNNDFSKAEETLKALKDSVKDLGINPEWISKI
                     ENLNAALNDFKNGKNKDFSKVTQAKSDLENSIKDVIINQKITDKVDNLNQAVSEIKLT
                     GDFSKVEQALAELKNLSLDLGKNSDLQKSVKNGVNGTLVSNGLSKTEATTLTKNFSDI
```



```
                     RKELNEKLFGNSNNNNGLKNNTEPIYAQVNKKKTGQATSPEEPIYAQVAKKVSAKID
                     QLNEATSAINRKIDRINKIASAGKGVGGFSGAGRSASPEPIYATIDPDEANQAGFPLR
                     RSAAVNDLSKVGLSREQELTRRIGDLSQAVSEAKTGHFGNLEQKIDELKDSTKKNALK
                     LWVESAKQVPTELQAKLDNYATNSHTRINSNVQSGTINEKATGMLTQKNPEWLKLVND
                     KIVAHNVGSAPLSAYDKIGFNQKNMKDYSDSFKFSTKLNNAVKDIKSSFVQPLTNTFS
                     TGSYSLMKANVEHGVKNTNTKGGFQKS"
BASE COUNT     1321 a    614 c    707 g    880 t
ORIGIN


  Query Match             1.6%;  Score 94;  DB 1;  Length 3522;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-33;
  Matches   94;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2326 gctaaaaacacaggcaattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatct 2385
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1795 GCTAAAAACACAGGCAATTATGATGAAGTGAAAAAAGCTCAGAAAGATCTTGAAAAATCT 1854

Qy   2386 ctaaggaaacgagagcatttagagaaagaagtag 2419
           |||||||||||||||||||||||||||||||||
Db   1855 CTAAGGAAACGAGAGCATTTAGAGAAAGAAGTAG 1888


RESULT 118
HPY239723
LOCUS       HPY239723     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain BO265.
ACCESSION   AJ239723
VERSION     AJ239723.1  GI:4467764
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1..243
                     /organism="Helicobacter pylori"
                     /strain="Bo265"
```



```
                     /db_xref="taxon:210"
                     /country="Germany"
     gene            1..243
                     /gene="cagA"
     CDS             <1..>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37015.1"
                     /db_xref="GI:4467765"
                     /db_xref="SPTREMBL:O9Z5M2"
                     /translation="VGYKDQQGNNVATIINMHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIQNKVDFMEFLAQNNAKLDNLSE"
BASE COUNT       96 a     40 c     49 g     58 t
ORIGIN


  Query Match             1.6%;  Score 92;  DB 2;  Length 243;
  Best Local Similarity   99.0%;  Pred. No. 2.4e-32;
  Matches  192;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   1586 tgcatatgaaaaacggcagtggcttagtcatagcaggtggtgagaaagggattaacaacc 1645
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     50 TGCATATGAAAAACGGCAGTGGCTTAGTCATAGCAGGTGGTGAGAAAGGGATTAACAACC 109

Qy   1646 ctagtttttatctctacaaagaagaccaactcacaggctcacaacgagcattaagtcaag 1705
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    110 CTAGTTTTTATCTTTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATTAAGTCAAG 169

Qy   1706 aagagatccaaaacaaaatagatttcatggaatttcttgcacaaaataatgctaaattag 1765
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    170 AAGAGATCCAAAACAAAGTAGATTTCATGGAATTTCTTGCACAAAATAATGCTAAATTAG 229

Qy   1766 acaacttgagcgag 1779
           ||||||||||||||
Db    230 ACAACTTGAGCGAG 243


RESULT 119
AF202224
LOCUS       AF202224      286 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain Indial8A CagA (cagA) gene, partial cds.
ACCESSION   AF202224
VERSION     AF202224.1  GI:7229439
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 286)
```

```
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
            Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
            Nair,G.B. and Berg,D.E.
  TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
            India
  JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
  MEDLINE   20270153
   PUBMED   10809703
REFERENCE   2  (bases 1 to 286)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Chowdhury,A., Datta,S., Nair,G.B.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) Molecular Microbiology, Washington
            University, 660 South Euclid Ave., St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .286
                     /organism="Helicobacter pylori"
                     /strain="India18A"
                     /db_xref="taxon:210"
     gene            <1. .>286
                     /gene="cagA"
     CDS             <1. .>286
                     /gene="cagA"
                     /note="similar to Helicobacter pylori 26695 HP0547"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA"
                     /protein_id="AAF42826.1"
                     /db_xref="GI:7229440"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT      105 a      52 c      66 g      63 t
ORIGIN


  Query Match           1.5%;  Score 89;  DB 1;  Length 286;
  Best Local Similarity  100.0%;  Pred. No. 6.4e-31;
  Matches   89;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1609 ttagtcatagcaggtggtgagaaagggattaacaacccctagtttttatctctacaaagaa 1668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    115 TTAGTCATAGCAGGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAA 174

Qy   1669 gaccaactcacaggctcacaacgagcatt 1697
          |||||||||||||||||||||||||||||
Db    175 GACCAACTCACAGGCTCACAACGAGCATT 203


RESULT 120
AF289401
LOCUS       AF289401      291 bp    DNA             BCT       02-SEP-2000
```



```
DEFINITION  Helicobacter pylori isolate CR4 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289401
VERSION     AF289401.1  GI:9964156
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR4"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09850.1"
                     /db_xref="GI:9964157"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YQIFTSWVSHQNDPSKINTRSIRNFMEHIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      107 a      62 c      45 g      77 t
ORIGIN


  Query Match           1.5%;  Score 89;  DB 2;  Length 291;
  Best Local Similarity  100.0%;  Pred. No. 6.4e-31;
```



```
  Matches   89;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 60

Qy    811 ttaatcaacaaagacaatctcattgatgt 839
          |||||||||||||||||||||||||||||
Db     61 TTAATCAACAAAGACAATCTCATTGATGT 89


RESULT 121
AF289412
LOCUS       AF289412      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR22 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289412
VERSION     AF289412.1  GI:9964178
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR22"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
```



```
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09861.1"
                     /db_xref="GI:9964179"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YQIFTSWVSHQNDPSKINTRSIRNFMEHIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      107 a      62 c      45 g      77 t
ORIGIN


  Query Match           1.5%;  Score 89;  DB 2;  Length 291;
  Best Local Similarity  100.0%;  Pred. No. 6.4e-31;
  Matches   89;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 60

Qy    811 ttaatcaacaaagacaatctcattgatgt 839
          |||||||||||||||||||||||||||||
Db     61 TTAATCAACAAAGACAATCTCATTGATGT 89


RESULT 122
AF289414
LOCUS       AF289414      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR29 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289414
VERSION     AF289414.1  GI:9964182
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
```

```
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR29"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09863.1"
                     /db_xref="GI:9964183"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPSKINTRSIRIFMEHIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      105 a     62 c     46 g     78 t
ORIGIN


  Query Match             1.5%;  Score 89;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-31;
  Matches   89;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 60

Qy     811 ttaatcaacaaagacaatctcattgatgt 839
           |||||||||||||||||||||||||||||
Db      61 TTAATCAACAAAGACAATCTCATTGATGT 89


RESULT 123
AF289423
LOCUS       AF289423      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR52 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289423
VERSION     AF289423.1  GI:9964200
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
```



```
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR52"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09872.1"
                     /db_xref="GI:9964201"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPPKINTRSIRNFMENIIQPPIHDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      108 a     58 c     46 g     79 t
ORIGIN


  Query Match             1.5%;  Score 89;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-31;
  Matches   89;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     751 atcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgat 810
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGAT 60

Qy     811 ttaatcaacaaagacaatctcattgatgt 839
           |||||||||||||||||||||||||||||
Db      61 TTAATCAACAAAGACAATCTCATTGATGT 89
```



```
RESULT 124
AF289426
LOCUS       AF289426      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR55 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289426
VERSION     AF289426.1  GI:9964206
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR55"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09875.1"
                     /db_xref="GI:9964207"
                     /translation="VKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIFTSWVSHQNDPYKINTRSIRNFMEHIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a     61 c     47 g     77 t
```



```
ORIGIN


  Query Match             1.5%;  Score 88;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-30;
  Matches   88;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     752 tcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaataagagcaatgatt 811
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 TCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATT 61

Qy     812 taatcaacaaagacaatctcattgatgt 839
           |||||||||||||||||||||||||||||
Db      62 TAATCAACAAAGACAATCTCATTGATGT 89


RESULT 125
AF289441
LOCUS       AF289441      678 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR13 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289441
VERSION     AF289441.1  GI:9964236
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Helicobacter pylori"
                     /isolate="CR13"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>678
                     /gene="cagA"
```

```
     CDS             <1. .>678
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09890.1"
                     /db_xref="GI:9964237"
                     /translation="NFSDIKKELNAKLFRNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYTQVAKKVNAKIDRLNQIASGLGGVGKAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDPVGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDD
                     LSKVGLSREQQLKQKIDNLNQAVSEAKAGYFGNLEQTIDKLKDSTKHNPMNLWVESAK
                     KVPANLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      243 a    121 c    147 g    167 t
ORIGIN


  Query Match             1.5%;  Score 88;  DB 2;  Length 678;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-30;
  Matches   88;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3490 ctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatgaatctatgg 3549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 CTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATGAATCTATGG 639

Qy   3550 gttgaaagtgcaaaaaaagtacctgcta 3577
          |||||||||||||||||||||||||||
Db    640 GTTGAAAGTGCAAAAAAAGTACCTGCTA 667


RESULT 126
AE000560/c
LOCUS       AE000560    10351 bp    DNA             BCT       06-APR-1999
DEFINITION  Helicobacter pylori 26695 section 38 of 134 of the complete genome.
ACCESSION   AE000560 AE000511
VERSION     AE000560.1  GI:2313554
KEYWORDS    .
SOURCE      Helicobacter pylori 26695.
  ORGANISM  Helicobacter pylori 26695
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 10351)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
```



```
            Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and
            Venter,J.C.
  TITLE     The complete genome sequence of the gastric pathogen Helicobacter
            pylori
  JOURNAL   Nature 388 (6642), 539-547 (1997)
  MEDLINE   97394467
  REMARK    Erratum:[[published erratum appears in Nature 1997 Sep
            25;389(6649):412]]
REFERENCE   2  (bases 1 to 10351)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C.,
            Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M.,
            Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 10351)
  AUTHORS   White,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10351
                     /organism="Helicobacter pylori 26695"
                     /strain="26695"
                     /db_xref="taxon:85962"
     gene            complement(88. .273)
                     /gene="HP0445"
     CDS             complement(88. .273)
                     /gene="HP0445"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0445"
                     /protein_id="AAD07521.1"
                     /db_xref="GI:2313562"
                     /translation="MQPPLKKDLRISQKYVIPKGFDPTNKEYKNFLVKDLPKLVHREV
                     EIVMKLLFEINEVVDRD"
     gene            complement(292. .1071)
                     /gene="HP0446"
     CDS             complement(292. .1071)
                     /gene="HP0446"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
```



```
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0446"
                     /protein_id="AAD07516.1"
                     /db_xref="GI:2313557"
                     /translation="MQERVFKRKVLDANILKEMHANNVCYSKHSKDRFIPFKFDKFGY
                     VGCKLFKKILNFPSNTTFFGGTGCKKLMELLSEIVIDSRSSKIALNRHYALTRLQWCD
                     RTLRHNLQILERIGFLTAFKNKKGYIPLSMHDFTKIENYEHSGLNGESNLPNSFFLGI
                     CGYLKKLFKKLKDRAFRLANKHGVFFLKIPKHFQMQNFNNIFLEFVSVNNPCFSYRLT
                     YDQLVGKKIPNIKCSYQQAIVKKNIHRALDELSIDKEILAS"
     gene            1179. .2261
                     /gene="HP0447"
     CDS             1179. .2261
                     /gene="HP0447"
                     /note="similar to SP:P34243 PID:395256 PID:486007 percent
                     identity: 38.22; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD07514.1"
                     /db_xref="GI:2313555"
                     /translation="MSLANDLLEILSSQSPNDKRIRVLLEYISVLERTPWDLESLLRD
                     YNFVFSNTTGQHNQALERKETPYFDSVIVDEAAKANPLELLMVMALAKERIILVGDDR
                     QLPHYLDDEIGKKLEDESQDAQDEIEKALKDSMFKKLKERAQKLKELDGKECFITLNM
                     QYGMHPLLGELVSDVFYKPHNESFESPLKGKHLEEKPFKHNLRVLDNKPLAWIDVKES
                     KEKRNADGSYYRESEITAIKECLDLFMKDEPDFTPGVITFPSEQKRLLEQALKGYANL
                     EIGTVDSFQGKEFDVVFLSSVRTRHTEGFGFLKISSCLCVALSRQKRALIVAGDREKF
                     DTPESKDKVSGLFEFLQSCKKEGKIL"
     gene            2258. .2644
                     /gene="HP0448"
     CDS             2258. .2644
                     /gene="HP0448"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0448"
                     /protein_id="AAD07517.1"
                     /db_xref="GI:2313558"
                     /translation="MKIISFENDSVCDKHLLIPCRIYCVELEIRHNDLGLNAIEKCAL
                     KLKESGVKDEELKGFLGPGGELDLLEPILEKIETKTLEEYKKIHAHFYYNLINQEFLH
                     FVDLERVRAIDGKEVKCPTAKDDWAS"
     gene            2891. .4309
                     /gene="HP0449"
     CDS             2891. .4309
                     /gene="HP0449"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0449"
```



```
                     /protein_id="AAD07518.1"
                     /db_xref="GI:2313559"
                     /translation="MEFPADSSGFKAESLYYPKENNKSLNVNEIIIRKTIETSIKYHKQ
                     DTLRHAKVVSWHVLGEVYYLHAKAFDDSKEIRVECKNHPIEKMAEKLEETNPEWFEKW
                     REKQYTQTGESKPSKRIKVFKNFTAFDDRLYTIECNLKNLDTHQKKFEICGALYDIYE
                     QIFDETPSLKGRDLETYKAQDLSKKFMHLGFEQISKDLNDSRLNALLCYEEKVMQALA
                     KKYPSPLQDLHDIKKYRNKDKHGEKPQDGSSLTRVELERYRDGIYFLVENLLKNPLIK
                     ERENAQEEKHYKKNAEIDDRSQLSNLNAPKPLFECFVGVNLAKAKYYSKKEEREKEKM
                     ILNFCKIFEIILFEAIQKQPKDFKNKDELLGDYPNLKNLDSLREVREDFLKRAFKND
                     EASLGAYVLVLLSCKYFESVFEKVQEWLDFIARLIALRGHVHKITKELERLEEEDLEK
                     LEKQALEYFNKIANKIYLKEKR"
     gene            4306. .4440
                     /gene="HP0450"
     CDS             4306. .4440
                     /gene="HP0450"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0450"
                     /protein_id="AAD07522.1"
                     /db_xref="GI:2313563"
                     /translation="MSGNEELELRARETELDKRVAELDKREKQLRKDIDAFKQEKKFI
                     "
     gene            4551. .4835
                     /gene="HP0451"
     CDS             4551. .4835
                     /gene="HP0451"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0451"
                     /protein_id="AAD24378.1"
                     /db_xref="GI:4580426"
                     /translation="MILETISKQLQEQQESLNKDYRKKLESIKERSDRLESDLKTQFD
                     AQLEKWEQNFKGYEQQLTDILNERELQLDLDQQRLNAQKQALENHMKQREE"
     gene            4911. .6467
                     /gene="HP0452"
     CDS             4911. .6467
                     /gene="HP0452"
                     /note="similar to GP:2983256 percent identity: 34.34;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD07515.1"
                     /db_xref="GI:2313556"
                     /translation="MADETNALRQKNRELNKKIDWQKDYDREKLERDNRDLEQCKEDL
                     LAANENLRNRIQEWENEKNKLDPRDERIKELEEEKRELEGILAQKENAEQKYNTLSVK
                     NKQLEAELDMLNEKFEKLKNMYAGVEDPEKRQKNIKEQIVKTNPKVLGAPSNEVEELA
```

```
                     FLERIEKGMQEFNVFYPKRLLYMPHTALKSTSLSPLSVLSGVSGTGKSELPKLYVHFG
                     GLNFLSIAVQPTWDSPESLMGYFHAIENKFDATEPLRFFIQTTLSNNEEPYGLKEAMS
                     IVLLDEMNLAHIELYFAEFLSKLEIKCSQETNISIKLGTGLTWELPLGDNLLFVGTMN
                     EDESTKMLSDKVLDRAFCLNFERPKILKGKQQKPIPSNDGYLKVETFNRWINKKGDQE
                     AKLEGKYKKLTEEINERLDACCRSIGHRVWQSMSTYMHNHPLVLHASDKNKALQFAYE
                     ECLVLKIFPKLRGVQTRNNQHLTKIQDLLKDFSVSSDFKQAMENDFKEFVFNSTNYLN
                     NVEYEKLLKK"
     gene            6474. .9539
                     /gene="HP0453"
     CDS             6474. .9539
                     /gene="HP0453"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0453"
                     /protein_id="AAD07519.1"
                     /db_xref="GI:2313560"
                     /translation="MDLEELYAPNHIERLKARSPLRSIAFFDDFSASFEYRDLFSVLE
                     NIVQFDYEKKPYKDDLYFLCKFVEPALKAIFSNLNTNIYRKHLKMPLEKAREFDAKCA
                     LDLAKRPGRSLKEKLCDNKVLSVKRYVNANTHENRFLKRFIKELLRIIHWREIEFQQV
                     PEELIFSITSFLKNGVAQQIDEKQAIIPNNLLHFDKHYKRIFKAHDWLYDGVGSLMNL
                     DQIFYLECLYQAQFYTSKNIEPTLIRNEQDLYALIKNSFPIKDLSFEKMRLKAKEFFE
                     NELRQPINLDQEIPQLELCKGVYKEMYIDMFSPEPFALLVGNGNEEKILKLPLLVKKQ
                     ENNTYINANGAKGKIDEKGYLANALKNYDETLVEAFMRDFKERYKIEKLYYLLDDNIK
                     NFEFMKIFHKISLYFKDAKFYPKSVALGFSSLFENKLKNERLRYNSVDLVVKENHKS
                     KTFNDCGLVLERQKSDDSKEFLILQDSFIKKALKNFKRALGLEKEGFILYKECIDPRLS
                     MEVVKDGRFKNFEIIKDKTILGDKETLEIETPFIIPKGRESFALPLILNEEKIAYQGK
                     ITSKDFPLENDEEYKLTLTYDIGTEFNYVLEFKPVNNDLKPIVMEWQRIDRVELPTPD
                     SIKKPSIDELKNDFNPKRGKSSDLPEWALEQLETLKDLNSPPRFVLEKKLECGGISII
                     GEDRNNELFYIMETNGKKVFCHSRQCKGSVNKDELSLGARVCLEVGPDKNDHGKYRGK

  Query Match             1.5%;  Score 88;  DB 1;  Length 10351;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-30;
  Matches   88;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4222 tagggacacagagtgagaaccaaaactctccctacgggcaacatcagcctaggaagcccaa 4281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2691 TAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCAACATCAGCCTAGGAAGCCCAA 2632

Qy   4282 tcgtctttagcggttgggcacttcacct 4309
          ||||||||||||||||||||||||||||
Db   2631 TCGTCTTTAGCGGTTGGGCACTTCACCT 2604


RESULT 127
AF289424
LOCUS       AF289424      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR53 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289424
```



```
VERSION     AF289424.1  GI:9964202
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR53"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09873.1"
                     /db_xref="GI:9964203"
                     /translation="IKNPTKKNQYFSDFINKSNDLINKDNLIDVESSTESFRKFGDQR
                     YRIPTSWVSHQNDPSKINTRSIRNFMEHTIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT      106 a      65 c      45 g      75 t
ORIGIN


  Query Match             1.5%;  Score 87;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-30;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    864 gaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatcc 923
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 GAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCC 173

Qy    924 gtctaaaatcaacacccgatcgatccg 950
          |||||||||||||||||||||||||||
Db    174 GTCTAAAATCAACACCCGATCGATCCG 200


RESULT 128
HPY239734
LOCUS       HPY239734     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain 5-1.
ACCESSION   AJ239734
VERSION     AJ239734.1  GI:4467786
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Helicobacter pylori"
                     /strain="5-1"
                     /db_xref="taxon:210"
                     /country="Lithuania"
     gene            1. .243
                     /gene="cagA"
     CDS             <1. .>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37826.1"
                     /db_xref="GI:4467787"
                     /db_xref="SPTREMBL:Q9Z5L4"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIRNKVDPMEFLAQNNAKLDNLSE"
BASE COUNT       93 a      41 c      52 g      57 t
```



```
ORIGIN


  Query Match             1.5%;  Score 86;  DB 2;  Length 243;
  Best Local Similarity   99.3%;  Pred. No. 1.7e-29;
  Matches  136;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaatacttaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     85 GGTGGTGAGAAAGGGATTAATAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    145 GGCTCACAACGAGCATT 161


RESULT 129
HPY239735
LOCUS       HPY239735     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain 96-228.
ACCESSION   AJ239735
VERSION     AJ239735.1  GI:4467788
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Helicobacter pylori"
                     /strain="96-228"
                     /db_xref="taxon:210"
                     /country="USA:Alaska"
     gene            1. .243
```

```
                     /gene="cagA"
      CDS            <1. .>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37827.1"
                     /db_xref="GI:4467789"
                     /db_xref="SPTREMBL:Q9Z5L3"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNFSE"
BASE COUNT       93 a      43 c      51 g      56 t
ORIGIN


  Query Match             1.5%;  Score 86;  DB 2;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-29;
  Matches   86;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1612 gtcatagcaggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagac 1671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 GTCATAGCAGGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGAC 135

Qy   1672 caactcacaggctcacaacgagcatt 1697
          |||||||||||||||||||||||||
Db    136 CAACTCACAGGCTCACAACGAGCATT 161


RESULT 130
HPY252963
LOCUS       HPY252963     244 bp    DNA             BCT       08-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
            protein A.
ACCESSION   AJ252963
VERSION     AJ252963.1  GI:5851957
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
REFERENCE   2  (bases 1 to 244)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) van der Ende A., Medical Microbiology,
```



```
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES              Location/Qualifiers
     source           1. .244
                      /organism="Helicobacter pylori"
                      /strain="Dutch 107"
                      /db_xref="taxon:210"
                      /country="Netherlands"
     gene             1. .244
                      /gene="cagA"
     CDS              <1. .>244
                      /gene="cagA"
                      /codon_start=1
                      /transl_table=11
                      /product="cytotoxin associated protein A"
                      /protein_id="CAB55439.1"
                      /db_xref="GI:5851958"
                      /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                      DQLTGSQRALSQEEIRNKMDFMEFLAQNNAKLDNLSE"
BASE COUNT       94 a      41 c      52 g      57 t
ORIGIN


  Query Match             1.5%;  Score 86;  DB 2;  Length 244;
  Best Local Similarity   99.3%;  Pred. No. 1.7e-29;
  Matches  136;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     85 GGTGGTGAGAAAGGGATTAATAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    145 GGCTCACAACGAGCATT 161


RESULT 131
HPY252964
LOCUS       HPY252964     244 bp    DNA             BCT       08-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
            protein A.
ACCESSION   AJ252964
VERSION     AJ252964.1  GI:5851959
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
```



```
            Helicobacter.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
REFERENCE   2  (bases 1 to 244)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES              Location/Qualifiers
     source           1. .244
                      /organism="Helicobacter pylori"
                      /strain="Dutch 137"
                      /db_xref="taxon:210"
                      /country="Netherlands"
     gene             1. .244
                      /gene="cagA"
     CDS              <1. .>244
                      /gene="cagA"
                      /codon_start=1
                      /transl_table=11
                      /product="cytotoxin associated protein A"
                      /protein_id="CAB55440.1"
                      /db_xref="GI:5851960"
                      /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                      DQLTGSQRALSQEEIRNKVDFMEFLAQNNAKLDNLSE"
BASE COUNT       94 a      40 c      52 g      58 t
ORIGIN


  Query Match             1.5%;  Score 86;  DB 2;  Length 244;
  Best Local Similarity   99.3%;  Pred. No. 1.7e-29;
  Matches  136;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 84

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     85 GGTGGTGAGAAAGGGATTAACAATCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 144

Qy   1681 ggctcacaacgagcatt 1697
          |||||||||||||||||
Db    145 GGCTCACAACGAGCATT 161
```



```
RESULT 132
AF202223
LOCUS       AF202223      286 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain India17A CagA (cagA) gene, partial cds.
ACCESSION   AF202223
VERSION     AF202223.1  GI:7229437
KEYWORDS
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
            Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
            Nair,G.B. and Berg,D.E.
  TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
            India
  JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
  MEDLINE   20270153
   PUBMED   10809703
REFERENCE   2  (bases 1 to 286)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Chowdhury,A., Datta,S., Nair,G.B.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) Molecular Microbiology, Washington
            University, 660 South Euclid Ave., St. Louis, MO 63110, USA
FEATURES              Location/Qualifiers
     source           1. .286
                      /organism="Helicobacter pylori"
                      /strain="India17A"
                      /db_xref="taxon:210"
     gene             <1. .>286
                      /gene="cagA"
     CDS              <1. .>286
                      /gene="cagA"
                      /note="similar to Helicobacter pylori 26695 HP0547"
                      /codon_start=1
                      /transl_table=11
                      /product="CagA"
                      /protein_id="AAF42825.1"
                      /db_xref="GI:7229438"
                      /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                      EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT      102 a      53 c      68 g      63 t
ORIGIN


  Query Match             1.5%;  Score 86;  DB 1;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-29;
  Matches   86;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy   1612 gtcatagcaggtggtgagaaagggattaacaaccctagttttatctctacaaagaagac 1671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 GTCATAGCAGGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGAC 177

Qy   1672 caactcacaggctcacaacgagcatt 1697
          ||||||||||||||||||||||||||
Db    178 CAACTCACAGGCTCACAACGAGCATT 203


RESULT 133
AF191013
LOCUS       AF191013      794 bp    DNA              BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain India29A cag pathogenicity island
            polymorphic right end, type Ia motif.
ACCESSION   AF191013
VERSION     AF191013.1  GI:10567576
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 794)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 794)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .794
                     /organism="Helicobacter pylori"
                     /strain="India29A"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="India: Calcutta"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ia motif; similar to
                     Helicobacter pylori strain Gambia94/24 type Ia motif
```



```
                     presented in GenBank Accession Number AF190658; contains
                     46 bp deletion compared to Helicobacter pylori strain
                     NCTC11638"
BASE COUNT      263 a    133 c    144 g    254 t
ORIGIN

  Query Match             1.5%;  Score 86;  DB 1;  Length 794;
  Best Local Similarity   99.3%;  Pred. No. 1.6e-29;
  Matches  136;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   4557 caaagaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaaagacgc 4616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    267 CAAAGAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAAGATGC 326

Qy   4617 ttgcgatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacact 4676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 TTGCGATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACT 386

Qy   4677 gtggatgcctttcaagg 4693
          |||||||||||||||||
Db     387 GTGGATGCCTTTCAAGG 403


RESULT 134
AF125214/c
LOCUS       AF125214      2693 bp    DNA              BCT       26-APR-1999
DEFINITION  Helicobacter pylori cag pathogenicity island CagA (cagA) gene,
            partial cds; and urease-enhancing factor Lpp (lpp),
            urease-enhancing factor helB (helB), and HelA (helA) genes,
            complete cds.
ACCESSION   AF125214
VERSION     AF125214.1  GI:4680600
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 2693)
  AUTHORS   McGee,D.J., May,C.A., Garner,R.M., Himpsl,J.M. and Mobley,H.L.T.
  TITLE     Isolation of Helicobacter pylori genes that modulate urease
            activity
  JOURNAL   J. Bacteriol. 181 (8), 2477-2484 (1999)
  MEDLINE   99214098
REFERENCE   2  (bases 1 to 2693)
  AUTHORS   McGee,D.J., May,C.A., Garner,R.M., Himpsl,J.M. and Mobley,H.L.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-1999) Microbiology & Immunology, University of
            Maryland, Baltimore, 655 W. Baltimore St., Baltimore, MD 21201, USA
FEATURES             Location/Qualifiers
     source          1. .2693
```



```
                     /organism="Helicobacter pylori"
                     /strain="ATCC 43504"
                     /db_xref="taxon:210"
                     /db_xref="ATCC:43504"
     gene            complement(<1. .563)
                     /gene="cagA"
     CDS             complement(<1. .563)
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cag pathogenicity island CagA"
                     /protein_id="AAD27691.1"
                     /db_xref="GI:4680601"
                     /translation="MTNETINQQPQTEAAFNPQQCINNLQVAFLKADHAVASYDPDQK
                     PIVDKNDRDNRQAFDGISQLREEYSNKAIKNPTKKNQYFSDWINKSKDLIHKDNVIKI
                     GYSKKSPQKFETQRYRIFTSWASHQNDPYKINTRSIRNSIDHIIQPPIPDDKEKAEFL
                     KSAKQSFAGIIIGNQIRTDQKFMGVLD"
     gene            951. .1139
                     /gene="lpp"
     CDS             951. .1139
                     /gene="lpp"
                     /note="hypothetical lipoprotein"
                     /codon_start=1
                     /transl_table=11
                     /product="urease-enhancing factor Lpp"
                     /protein_id="AAD27692.1"
                     /db_xref="GI:4680602"
                     /translation="MKTIRNSMFIGASLLGGCASVETRFDALRVACVKDAVIEKESHH
                     TPKDFGSFYFSNQTDIGL"
     gene            complement(1416. .1973)
                     /gene="helB"
     CDS             complement(1416. .1973)
                     /gene="helB"
                     /note="putative DNA helicase"
                     /codon_start=1
                     /transl_table=11
                     /product="urease-enhancing factor helB"
                     /protein_id="AAD27693.1"
                     /db_xref="GI:4680603"
                     /translation="MPASIGSLVSQLFYKEKLKNGVIKNTSQFYDPKNIIRWINVEGE
                     HKLEKTSSYNKNQVQKIIELLEQINRILNQRKIRKTIGIITPYNAQKRCLRSEVEKYG
                     FKNFDELKIDTVDAFQGEKADIIIYSTVKTYGNLSFLIDSKRLNVLFLGQKKISFLWA
                     KSLSLRICEAMERISLALFCKSVDR"
     gene            complement(1980. .2246)
                     /gene="helA"
     CDS             complement(1980. .2246)
                     /gene="helA"
                     /note="putative DNA helicase"
                     /codon_start=1
                     /transl_table=11
                     /product="HelA"
```



```
                     /protein_id="AAD27694.1"
                     /db_xref="GI:4680604"
                     /translation="MEFDVTIIDETGRATAPEILIPALRTKKLILIGDHNQLPPSIDR
                     YLLEQLESDDIQNLDAIDRQLLEESFFENLYKYIPESNKAMLNE"
     repeat_region   complement(2378. .>2693)
                     /note="mini IS605"
                     /rpt_family="insertion sequence IS605"
                     /rpt_type=dispersed
BASE COUNT      791 a    467 c    487 g    948 t
ORIGIN

  Query Match             1.5%;  Score 86;  DB 1;  Length 2693;
  Best Local Similarity   99.3%;  Pred. No. 1.5e-29;
  Matches  136;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   4557 caaagaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaaagacgc 4616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   1754 CAAAGAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAAGATGC 1695

Qy   4617 ttgcgatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacact 4676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1694 TTGCGATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACT 1635

Qy   4677 gtggatgcctttcaagg 4693
          |||||||||||||||||
Db   1634 GTGGATGCCTTTCAAGG 1618


RESULT 135
AF289458
LOCUS       AF289458      567 bp    DNA              BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR54 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289458
VERSION     AF289458.1  GI:9964270
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Helicobacter pylori"
                     /isolate="CR54"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>567
                     /gene="cagA"
     CDS             <1. .>567
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09907.1"
                     /db_xref="GI:9964271"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPFPLKRHDKVDDLSKVGLSREQQLKQKIDNLSQAVSEAKAGFPGNL
                     EQTIDKLKDSTKNNTVNLWAESAKKVPAS"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      210 a     97 c    123 g    137 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 567;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 216

Qy   3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 276

Qy   3391 ctcagtaaggtagggc 3406
          ||||||||||||||||
Db    277 CTCAGTAAGGTAGGGC 292


RESULT 136
AF289439
```



```
LOCUS       AF289439      570 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR9 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289439
VERSION     AF289439.1  GI:9964232
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 570)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Helicobacter pylori"
                     /isolate="CR9"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>570
                     /gene="cagA"
     CDS             <1. .>570
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09888.1"
                     /db_xref="GI:9964233"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNEPIYAKVNKKKTGQVAS
                     PEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSPEP
                     IYATIDDLGGPFPLKRHDKVDDLSKVGLSREQKLAQKIDNLNQAVSEAKAGYFPGNLEQ
                     TIDKLKDSTKHNPMNLWVESAKKVPASLSA"
BASE COUNT      206 a     101 c    125 g    138 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 570;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
```



```
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    151 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 210

Qy   3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    211 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 270

Qy   3391 ctcagtaaggtagggc 3406
          ||||||||||||||||
Db    271 CTCAGTAAGGTAGGGC 286


RESULT 137
AF289447
LOCUS       AF289447      576 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR29 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289447
VERSION     AF289447.1  GI:9964248
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Helicobacter pylori"
                     /isolate="CR29"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>576
                     /gene="cagA"
     CDS             <1. .>576
```



```
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09896.1"
                     /db_xref="GI:9964249"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPFPLKRHDKVDDLSKVGLSREQQLKQKIDNLSQAVSEAKAGFFGNL
                     EQTIDKLKDSTKNNTVNLWAESAKKVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      211 a     99 c    126 g    140 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 576;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 216

Qy   3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 276

Qy   3391 ctcagtaaggtagggc 3406
          ||||||||||||||||
Db    277 CTCAGTAAGGTAGGGC 292


RESULT 138
AF289461
LOCUS       AF289461      576 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR57 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289461
VERSION     AF289461.1  GI:9964276
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
```

```
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Helicobacter pylori"
                     /isolate="CR57"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>576
                     /gene="cagA"
     CDS             <1. .>576
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09910.1"
                     /db_xref="GI:9964277"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNWLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPPPLKRHDKVDDLSKVGLSRGATIEQKIDNLSQAVSEAKAGFFGNL
                     EQTIDKLKDSTKNNTVNLWAESAKKVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      210 a      99 c     126 g     141 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 576;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 216

Qy   3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 276

Qy   3391 ctcagtaaggtagggc 3406
          ||||||||||||||||
Db    277 CTCAGTAAGGTAGGGC 292


RESULT 139
AF289463
LOCUS       AF289463      576 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR59 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289463
VERSION     AF289463.1  GI:9964280
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Helicobacter pylori"
                     /isolate="CR59"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>576
                     /gene="cagA"
     CDS             <1. .>576
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09912.1"
                     /db_xref="GI:9964281"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPPPLKRHDKVDNLSKVGLSREQQLKQKIDNLNQAVSEAKAGFFGNL
                     EQTIDKLKDSTKNNTVNLWAESAKKVPAGLSA"
```



```
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      212 a      99 c     125 g     140 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 576;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 216

Qy   3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 276

Qy   3391 ctcagtaaggtagggc 3406
          ||||||||||||||||
Db    277 CTCAGTAAGGTAGGGC 292


RESULT 140
AF289453
LOCUS       AF289453      633 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR49 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289453
VERSION     AF289453.1  GI:9964260
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 633)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .633
```



```
                     /organism="Helicobacter pylori"
                     /isolate="CR49"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>633
                     /gene="cagA"
     CDS             <1. .>633
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09902.1"
                     /db_xref="GI:9964261"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAKVNKKKTGQVASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFP
                     LKRHDKVDDLSKVGRSVSPEPIYATIDDLRWTFPPEKHDKVDDLSKVGLSREQELAQK
                     IDNLNQAVSEAKAGYFGNLEQTIDKLKDSTKHNPMNLWVESVKKVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      235 a      112 c     135 g     151 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 633;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    214 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 273

Qy   3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    274 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 333

Qy   3391 ctcagtaaggtagggc 3406
          ||||||||||||||||
Db    334 CTCAGTAAGGTAGGGC 349


RESULT 141
AF289455
LOCUS       AF289455      633 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR51 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289455
VERSION     AF289455.1  GI:9964264
```

```
KEYWORDS
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 633)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Helicobacter pylori"
                     /isolate="CR51"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>633
                     /gene="cagA"
     CDS             <1. .>633
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09904.1"
                     /db_xref="GI:9964265"
                     /translation="NFSDIKKELNAKLPGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAKVNKKKTGQVASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFP
                     LKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVGLSREQELAQK
                     IDNLNQAVSEAKAGYPGNLEQTIDKLKDSTKHNPMNLWVESVKKVPAVLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      234 a    110 c    136 g    153 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 633;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```



```
Qy    3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     214 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 273

Qy    3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     274 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 333

Qy    3391 ctcagtaaggtagggc 3406
           ||||||||||||||||
Db     334 CTCAGTAAGGTAGGGC 349


RESULT 142
AF289434
LOCUS       AF289434      672 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR4 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289434
VERSION     AF289434.1  GI:9964222
KEYWORDS
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 672)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Helicobacter pylori"
                     /isolate="CR4"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>672
                     /gene="cagA"
     CDS             <1. .>672
                     /gene="cagA"
                     /note="CagA"
```



```
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09883.1"
                     /db_xref="GI:9964223"
                     /translation="NFSDIKKELNAKLPGNFNNNNNNGLKNEPIYAKVNKKKAGQVAS
                     PEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSPEP
                     IYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLS
                     KVGLSREQKLAQKIDNLNQAVSEAKAGYPGNLEQTIDKLKDSAKNNPVNLWVESAKKV
                     PANLSA"
BASE COUNT      234 a    118 c    150 g    170 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 672;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 210

Qy    3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 270

Qy    3391 ctcagtaaggtagggc 3406
           ||||||||||||||||
Db     271 CTCAGTAAGGTAGGGC 286


RESULT 143
AF289451
LOCUS       AF289451      672 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR43 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289451
VERSION     AF289451.1  GI:9964256
KEYWORDS
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 672)
```



```
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Helicobacter pylori"
                     /isolate="CR43"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>672
                     /gene="cagA"
     CDS             <1. .>672
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09900.1"
                     /db_xref="GI:9962577"
                     /translation="NFSDIKKELNAKLPGNFNNNNNNGLKNEPIYAKVNKKKTGQVAS
                     PEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSPEP
                     IYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLS
                     KVGLSREQQLAQKIDNLSQAVSEAKAGFFGNLEQTIDKLKDSTKHNPMNLWVESAKKV
                     PAILSA"
BASE COUNT      232 a    120 c    148 g    172 t
ORIGIN


  Query Match             1.4%;  Score 85;  DB 2;  Length 672;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-29;
  Matches  135;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 210

Qy    3331 ttgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgat 3390
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 TTGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGAT 270

Qy    3391 ctcagtaaggtagggc 3406
           ||||||||||||||||
Db     271 CTCAGTAAGGTAGGGC 286


RESULT 144
AF289443
```

```
LOCUS       AF289443      570 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR20 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289443
VERSION     AF289443.1  GI:9964240
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 570)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Helicobacter pylori"
                     /isolate="CR20"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>570
                     /gene="cagA"
     CDS             <1. .>570
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09892.1"
                     /db_xref="GI:9964241"
                     /translation="NPSDIKKELNEKFKNFNNNNNNGLKNSTEPIYAKVNKKKTGQVAS
                     PEEPIYTQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSPEP
                     IYATIDDLGGPFPLKRHDKVDDLSKVGLSREQKLAQKIDNLNQAVSEAKAGYFGNLEQ
                     TIDKLKDSTKHNPLNLWVESAKKVPASLSA"
     misc_feature    73. .96
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      212 a     99 c    121 g    138 t
ORIGIN
```



```
  Query Match             1.4%;  Score 84;  DB 2;  Length 570;
  Best Local Similarity   99.3%;  Pred. No. 1.5e-28;
  Matches  134;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3272 ctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggtt 3331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 CTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGTT 211

Qy   3332 tgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgatc 3391
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 TGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGATC 271

Qy   3392 tcagtaaggtagggc 3406
          |||||||||||||||
Db    272 TCAGTAAGGTAGGGC 286


RESULT 145
AF289462
LOCUS       AF289462      573 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR58 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289462
VERSION     AF289462.1  GI:9964278
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 573)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 573)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .573
                     /organism="Helicobacter pylori"
                     /isolate="CR58"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
```



```
     gene            <1. .>573
                     /gene="cagA"
     CDS             <1. .>573
                     /gene="cagA"
                     /note="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09911.1"
                     /db_xref="GI:9964279"
                     /translation="NPSDIKKELNAKLGNFNNNNNNNGLKNSTEPIYAKVNKKKAGQAA
                     SPEEPIYTQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSPE
                     PIYATIDDLGGPFPLKRHDKVDDLSKVGLSREQELTQKIDNLNQAVSEAKAGFFGNLE
                     QTIDNLKDSTKHNTVNLWVESTKKVPASLSA"
     misc_feature    73. .96
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      208 a    101 c    126 g    138 t
ORIGIN


  Query Match             1.4%;  Score 84;  DB 2;  Length 573;
  Best Local Similarity   99.3%;  Pred. No. 1.5e-28;
  Matches  134;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3272 ctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggtt 3331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    155 CTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGTT 214

Qy   3332 tgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgatc 3391
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    215 TGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGATC 274

Qy   3392 tcagtaaggtagggc 3406
          |||||||||||||||
Db    275 TCAGTAAGGTAGGGC 289


RESULT 146
AF289449
LOCUS       AF289449      676 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR38 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289449
VERSION     AF289449.1  GI:9964252
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 676)
```



```
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 676)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Helicobacter pylori"
                     /isolate="CR38"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>676
                     /gene="cagA"
     CDS             <1. .>676
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09898.1"
                     /db_xref="GI:9964253"
                     /translation="NPSDIKKELNAKLFGNFNNNNNNGLENSAEPIYAKVNKKKTGQV
                     ASPEEPIYTQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDFGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDD
                     LSKVGLSRNQKLAWKIDNLNQAVSEAKAGFFGNLEQTIGKLKDSTKNNTVNLWVESAK
                     KPASLSA"
BASE COUNT      232 a    117 c    153 g    174 t
ORIGIN


  Query Match             1.4%;  Score 84;  DB 2;  Length 676;
  Best Local Similarity   99.3%;  Pred. No. 1.5e-28;
  Matches  134;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3272 ctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggtt 3331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 CTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGTT 217

Qy   3332 tgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgatc 3391
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 TGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGATC 277
```

```
Qy   3392 tcagtaaggtagggc 3406
          ||||||||||||||||
Db    278 TCAGTAAGGTAGGGC 292


RESULT 147
AF289432
LOCUS       AF289432      678 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR1 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289432
VERSION     AF289432.1  GI:9964218
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Helicobacter pylori"
                     /isolate="CR1"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>678
                     /gene="cagA"
     CDS             <1. .>678
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09881.1"
                     /db_xref="GI:9964219"
                     /translation="NFSNIKKELNAKLFGNFNNNNNNGLENSAEPIYAKVNKKKTGQV
                     ASPEEPIYTQVAKKVNAKIDRLNQIASGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDFGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDD
```



```
                     LSKVGLSRNQKLAWKIDNLNQAVSEAKAGFPGNLEQTIGKLKDSTKNNTVNLWVESAK
                     KVPASLSA"
BASE COUNT      232 a    117 c    155 g    174 t
ORIGIN


  Query Match           1.4%;  Score 84;  DB 2;  Length 678;
  Best Local Similarity   99.3%;  Pred. No. 1.5e-28;
  Matches  134;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3272 ctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggtt 3331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 CTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGTT 217

Qy   3332 tgggtgttgtagggcaagcagcgggcttccctttgaaaaggcatgataaagttgatgatc 3391
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 TGGGTGGTGTAGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGATC 277

Qy   3392 tcagtaaggtagggc 3406
          |||||||||||||||
Db    278 TCAGTAAGGTAGGGC 292


RESULT 148
AF289446
LOCUS       AF289446      669 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR23 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289446
VERSION     AF289446.1  GI:9964246
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 669)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Helicobacter pylori"
```



```
                     /isolate="CR23"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>669
                     /gene="cagA"
     CDS             <1. .>669
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09895.1"
                     /db_xref="GI:9964247"
                     /translation="NFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKAGQVASP
                     EEPIYAQVAKKVNAKIDRLNQIASGLGGVGQVAGFPLKRHDKVDDLSKVGRSVSPEPI
                     YATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSK
                     VGLSREQQLKQKIDNLSQAVSEAKAGFPGNLEQTIDKLKDSTKNNTVNLWAESAKKVP
                     ASLSA"
BASE COUNT      234 a    116 c    152 g    167 t
ORIGIN


  Query Match           1.4%;  Score 83;  DB 2;  Length 669;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-28;
  Matches   83;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3255 tgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaa 3314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 TGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAA 191

Qy   3315 tcaaatagcaagtggtttgggtg 3337
          |||||||||||||||||||||||
Db    192 TCAAATAGCAAGTGGTTTGGGTG 214


RESULT 149
AF198481
LOCUS       AF198481      220 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain PeruGJ8 cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198481
VERSION     AF198481.1  GI:7248688
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
```



```
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 220)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 220)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .220
                     /organism="Helicobacter pylori"
                     /strain="PeruGJ8"
                     /db_xref="taxon:210"
                     /note="European type; isolated from gastric juice sample
                     from native Peruvian"
     gene            <1. .>220
                     /gene="cagA"
     CDS             <1. .>220
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43621.1"
                     /db_xref="GI:7248689"
                     /translation="KDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQL
                     TGSQRALSQEEIRNKVDFMEFLAQNNAKL"
BASE COUNT       87 a     37 c     44 g     52 t
ORIGIN


  Query Match           1.4%;  Score 80;  DB 1;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-26;
  Matches   80;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     17 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 76

Qy   1621 ggtggtgagaaagggattaa 1640
          ||||||||||||||||||||
Db     77 GGTGGTGAGAAAGGGATTAA 96


RESULT 150
AF198470
LOCUS       AF198470      294 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain SAfrica19 cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198470
VERSION     AF198470.1  GI:7248666
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 294)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 294)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .294
```



```
                     /organism="Helicobacter pylori"
                     /strain="SAfrica19"
                     /db_xref="taxon:210"
                     /country="South Africa"
                     /note="European type"
     gene            <1. .>294
                     /gene="cagA"
     CDS             <1. .>294
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43610.1"
                     /db_xref="GI:7248667"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSEKE*
BASE COUNT      110 a      51 c      67 g      66 t
ORIGIN


  Query Match             1.4%;  Score 80;  DB 1;  Length 294;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-26;
  Matches   80;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 127

Qy   1621 ggtggtgagaaagggattaa 1640
          ||||||||||||||||||||
Db    128 GGTGGTGAGAAAGGGATTAA 147


RESULT 151
AF289460
LOCUS       AF289460      567 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR56 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289460
VERSION     AF289460.1  GI:9964274
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
```



```
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Helicobacter pylori"
                     /isolate="CR56"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>567
                     /gene="cagA"
     CDS             <1. .>567
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09909.1"
                     /db_xref="GI:9964275"
                     /translation="NFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKAGQVASP
                     EEPIYAQVAKKVNAKIDRLNQIASGLGGVGRAAGFPLKRHDKVDDLSKVGLSASPEPI
                     YATIDDLGGPFPLKRHDKVDDLSKVGLSRNQKLVQKIDNLSQAVSEAKAGFFGNLEQT
                     IDKLKDSTKNNTVNLWVESAKKVPASLSA*
BASE COUNT      206 a      97 c     124 g     140 t
ORIGIN


  Query Match             1.3%;  Score 78;  DB 2;  Length 567;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-25;
  Matches   78;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3255 tgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaa 3314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 TGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAA 191

Qy   3315 tcaaatagcaagtggttt 3332
          ||||||||||||||||||
Db    192 TCAAATAGCAAGTGGTTT 209


RESULT 152
HPU95957
LOCUS       HPU95957     2378 bp    DNA             BCT       01-FEB-1999
DEFINITION  Helicobacter pylori insertion sequence IS606 transposase homologs A
```



```
            (tnpA) and B (tnpB) genes, complete cds.
ACCESSION   U95957
VERSION     U95957.1  GI:2114469
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 2378)
  AUTHORS   Kersulyte,D., Akopyants,N.S., Clifton,S.W., Roe,B.A. and Berg,D.E.
  TITLE     Novel sequence organization and insertion specificity of IS605 and
            IS606: chimaeric transposable elements of Helicobacter pylori
  JOURNAL   Gene 223 (1-2), 175-186 (1998)
  MEDLINE   99077840
REFERENCE   2  (bases 1 to 2378)
  AUTHORS   Akopyants,N.S. and Berg D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-1997) Molecular Microbiology, Washington
            University School of Medicine, 660 South Euclid Ave., St.Louis, MO
            63110, USA
FEATURES             Location/Qualifiers
     source          1. .2378
                     /organism="Helicobacter pylori"
                     /strain="84-183"
                     /db_xref="taxon:210"
                     /map="near right end of the cag pathogenicity island in
                     strain 84-183; approximately 0.7 kb downstream of 3' end
                     of cagA gene"
     source          286. .2253
                     /organism="Helicobacter pylori"
                     /strain="84-183"
                     /db_xref="taxon:210"
                     /insertion_seq="IS606"
                     /note="similar to H. pylori transposable element IS605,
                     GenBank Accession Numbers AC000108 and U60177"
     gene            complement(417. .833)
                     /gene="tnpA"
     CDS             complement(417. .833)
                     /gene="tnpA"
                     /note="TnpA; IS200 transposase homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase homolog A"
                     /protein_id="AAD11513.1"
                     /db_xref="GI:2114470"
                     /translation="MKKIDDMRHGRHCVFLMHTHLVFVTKYRRKAFNKEVIDFLGSVF
                     AKVCKDFESELVEFDGESDHVHLLINYPPKVSVSKLVNSLKGVSSRLTRQHHFKSVEA
                     SLWGKHLWSPSYFAGSCGGAPLEMIKQYIQEQETPH"
     gene            870. .2198
                     /gene="tnpB"
     CDS             870. .2198
```

```
                     /gene="tnpB"
                     /note="TnpB; IS1341 transposase homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase homolog B"
                     /protein_id="AAD11514.1"
                     /db_xref="GI:2114471"
                     /translation="MKVNKGFKPRLYPTKEQQDKLQRCFFVYNQAYNIGLNLLQEQYE
                     TNKDSPPKERKWKKSSELDKAIKHHLNARGLSFSSVIAQQSRMNVERALKDAFKVKDR
                     GFPKFKSKSGAKQSFSWNNQGFSIKDSDEERFKIPTLMKMPLMMRMHRDFPPHSKVKQ
                     IVISWSHRKYPVSFCVEYEQDITPIKNPKNGVGLDLNILDIACSCGVNNHKKLTDFKQ
                     YPTDMKELLGIEIDEELDTKRLIPTYSKLYSLKKVSKKFRKLQRKQSRRVLKSKQNKT
                     KLGGNFYKTQKKLNQAFDKSSHQKTDRYHKITSELSKQFELVVVEDLQVKNMTKRAKL
                     KNVKQKSGLNQSILNTSFYQIISFLDYKQQHNGKLLVKVPPQYTSKTCHCCGNINHKL
                     KLNHRQYWCLECGYREHRDINAANNIISKGLSLFGVGNIHADFKEQSLSC"
BASE COUNT       822 a    456 c    399 g    701 t
ORIGIN


  Query Match           1.3%;  Score 77;  DB 2;  Length 2378;
  Best Local Similarity   99.2%;  Pred. No. 2.9e-25;
  Matches  127;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    4713 atttattccaccgtgaaaacttgtggtaatctttctttcttgctagattctaaacgcttg 4772
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 ATTTATTCCACCGTGAAAACTTGTGGTAATCTTTCTTTCTTGCTAGATTCTAAACGCTTG 70

Qy    4773 aatgtggctatttctagggcaaaagaaaatctcatttttgtgggtaaaaagtctttcttt 4832
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 AATGTAGCTATTTCTAGGGCAAAAGAAAATCTCATTTTTGTGGGTAAAAAGTCTTTCTTT 130

Qy    4833 gagaattt 4840
           ||||||||
Db     131 GAGAATTT 138


RESULT 153
AF289404
LOCUS       AF289404       291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR7 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289404
VERSION     AF289404.1  GI:9964162
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
```



```
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /isolate="CR7"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09853.1"
                     /db_xref="GI:9964163"
                     /translation="IKNPTKKNQYFSDFIHKSNDLINKDNLIDVESSTESFRKFGDQR
                     YQIFTSWVSHQNDPSKINTRSIRNFMEYIIQPPIPDDKEKAEFLKSAKQSFAG"
     misc_feature    115. .135
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif A"
BASE COUNT       106 a     62 c     45 g     78 t
ORIGIN


  Query Match           1.3%;  Score 76;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-25;
  Matches   76;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     888 aattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 AATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGAT 197

Qy     948 ccgaaatttttatggaa 963
           |||||||||||||||||
Db     198 CCGAAATTTTATGGAA 213


RESULT 154
```



```
AF289436
LOCUS       AF289436       570 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR6 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289436
VERSION     AF289436.1  GI:9964226
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 570)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Helicobacter pylori"
                     /isolate="CR6"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>570
                     /gene="cagA"
     CDS             <1. .>570
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09885.1"
                     /db_xref="GI:9964227"
                     /translation="MFSDIKKELNAKLGNFNNNNNGLKNSGEPIYAQVNKKKTGQVAS
                     PEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGPPLKKHGKVDDLSKVGLSASPEP
                     IYATIDDLGGPFPLKRHDKVDDLSKVGLSREQQLKQKIDNLSQAVSEAKAGFPGNLEQ
                     TIDKLRDSTKNNTVNLWAEGAKKVPASLSA"
     misc_feature    70. .93
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT       203 a    101 c    129 g    137 t
ORIGIN
```



```
  Query Match           1.3%;  Score 76;  DB 2;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-25;
  Matches   76;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3255 tgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaa 3314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     135 TGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAA 194

Qy    3315 tcaaatagcaagtggt 3330
           ||||||||||||||||
Db     195 TCAAATAGCAAGTGGT 210


RESULT 155
HPU80065
LOCUS       HPU80065       269 bp    DNA             BCT       22-OCT-1997
DEFINITION  Helicobacter pylori strain 184, cytotoxin-associated CagA (cagA)
            gene, partial cds.
ACCESSION   U80065
VERSION     U80065.1  GI:2558700
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 269)
  AUTHORS   Yang,J.C., Wang,T.H., Wang,H.J., Kuo,C.H., Wang,J.T. and Wang,W.C.
  TITLE     Genetic analysis of the cytotoxin-associated gene and the
            vacuolating toxin gene in Helicobacter pylori strains isolated from
            Taiwanese patients
  JOURNAL   Am. J. Gastroenterol. 92 (8), 1316-1321 (1997)
  MEDLINE   97405753
REFERENCE   2  (bases 1 to 269)
  AUTHORS   Wang,W.-C., Wang,H.-J. and Kuo,C.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1996) Life Science Department, National Tsing Hua
            University, Hsinchu 300, Taiwan
FEATURES             Location/Qualifiers
     source          1. .269
                     /organism="Helicobacter pylori"
                     /strain="184"
                     /db_xref="taxon:210"
     gene            1. .269
                     /gene="cagA"
     CDS             <1. .>269
                     /gene="cagA"
                     /codon_start=3
                     /transl_table=11
                     /product="cytotoxin-associated CagA"
```

```
                     /protein_id="AAB81560.1"
                     /db_xref="GI:2558701"
                     /translation="REEYANKAIKNPTKKNQYFSDFINKSNDLINKDNLIAVDSSVES
                     FRKFGDQRYQIFTSWVSLQKDPSKINTQQIQNFMENIIQPPISDD"
BASE COUNT      100 a      53 c      42 g      74 t
ORIGIN


  Query Match             1.2%;  Score 73;  DB 2;  Length 269;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-23;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    740 ccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaata 799
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16 CCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATA 75

Qy    800 agagcaatgattt 812
          |||||||||||||
Db     76 AGAGCAATGATTT 88


RESULT 156
HPY269866
LOCUS       HPY269866     452 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            JA214 from Japan.
ACCESSION   AJ269866
VERSION     AJ269866.1  GI:5921309
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 452)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Helicobacter pylori"
                     /isolate="JA214"
                     /db_xref="taxon:210"
                     /country="Japan"
```



```
     gene            1. .452
                     /gene="cagA"
     CDS             1. .>452
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56439.1"
                     /db_xref="GI:5921310"
                     /translation="MTSEPIDQTTAPDQTDFVPQRFINNLQVAFIEVDSAVASFDPDQ
                     KPIVDRIDRDNRQAFEKISQLREEYANKAIKNPTKKNQYFSDFINKSNDLINKDNLIA
                     VDSSVESSRKFGDQRYQIFTSWVSLQKDPSKINTQQIRNFMENIIQPP"
BASE COUNT      163 a      93 c      75 g     121 t
ORIGIN


  Query Match             1.2%;  Score 73;  DB 2;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-23;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    740 ccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagactttatcaata 799
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    209 CCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATA 268

Qy    800 agagcaatgattt 812
          |||||||||||||
Db    269 AGAGCAATGATTT 281


RESULT 157
AF190996/c
LOCUS       AF190996      590 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain OhioP46 cag pathogenicity island
            polymorphic right end, type Ic motif right-end.
ACCESSION   AF190996
VERSION     AF190996.1  GI:10567559
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
```



```
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 590)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Helicobacter pylori"
                     /strain="OhioP46"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="USA: Ohio"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ic motif right-end; similar to
                     Helicobacter pylori strain 84-183 type Ic motif presented
                     in GenBank Accession Number AF190660; truncated inside of
                     full copy of IS606"
BASE COUNT      175 a     102 c     127 g     186 t
ORIGIN


  Query Match             1.2%;  Score 73;  DB 1;  Length 590;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-23;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4332 acgaaaggctttgttcttttaaagtctgcatggatatttcctaccccaaaaagacttaacc 4391
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 ACGAAAGGCTTTGTTCTTTTAAAGTCTGCATGGATATTTCCTACCCCAAAAAGACTTAACC 203

Qy   4392 ctttgcttaaaat 4404
          |||||||||||||
Db    202 CTTTGCTTAAAAT 190


RESULT 158
AE000609
LOCUS       AE000609     11680 bp    DNA             BCT       06-APR-1999
DEFINITION  Helicobacter pylori 26695 section 87 of 134 of the complete genome.
ACCESSION   AE000609 AE000511
VERSION     AE000609.1  GI:2314150
KEYWORDS    .
SOURCE      Helicobacter pylori 26695.
  ORGANISM  Helicobacter pylori 26695
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 11680)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
```



```
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and
            Venter,J.C.
  TITLE     The complete genome sequence of the gastric pathogen Helicobacter
            pylori
  JOURNAL   Nature 388 (6642), 539-547 (1997)
  MEDLINE   97394467
  REMARK    Erratum:[[published erratum appears in Nature 1997 Sep
            25;389(6649):412]]
REFERENCE   2  (bases 1 to 11680)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C.,
            Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M.,
            Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 11680)
  AUTHORS   White,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11680
                     /organism="Helicobacter pylori 26695"
                     /strain="26695"
                     /db_xref="taxon:85962"
     gene            171. .485
                     /gene="HP1005"
     CDS             171. .485
                     /gene="HP1005"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP1005"
                     /protein_id="AAD08060.1"
                     /db_xref="GI:2314159"
                     /translation="MSNIITDYSQYNEKQLHNFLNSIEKQLLKAERDKNKAIKKIQEC
                     ELQERMIRQVLAQKHSQEKEPTPSLLNTIASKDDPEYDVSFGDFNDFLQIAKQERERH
                     NA"
     gene            complement(1036. .1569)
```

```
                     /gene="HP1006"
     CDS             complement(1036. .1569)
                     /gene="HP1006"
                     /note="similar to PID:1103907 SP:Q44346 percent identity:
                     27.33; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conjugal transfer protein (traG)"
                     /protein_id="AAD08052.1"
                     /db_xref="GI:2314151"
                     /translation="MVSLKESKKFEERVYLFLDEPVRFGKLPFLLEMPALSRSYGVVL
                     IFITQSNALIEKYYGREDARIVNSTVAYKIIFKMDDLEYAKQVSEEVGKMTRKTRSHS
                     TEKGGLITGGTSSIGKEAWDLLSACDIMNIDKDEVIVLVSGHKAKPLKLKANYYPKNK
                     ELLSRINWEVKPNEEVF"
     gene            complement(1617. .2944)
                     /gene="HP1007"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     PID:565287 percent identity: 34.34; identified by sequence
                     similarity; putative;transposase-like protein, PS3IS"
     gene            2982. .3398
                     /gene="HP1008"
     CDS             2982. .3398
                     /gene="HP1008"
                     /note="similar to GB:L25848 GB:X52093 PID:1177217
                     PID:439619 PID:1150644 percent identity: 33.91; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="IS200 insertion sequence from SARA17"
                     /protein_id="AAD08053.1"
                     /db_xref="GI:2314152"
                     /translation="MKKLEDMKHGRHCVFLMHVHFVFVTKYRRSAFNKEVIDPLGSVF
                     AKVCEDPESELVEPDGESDHVHLLINYPPKVSVSKLVNSLKGVSSRLTRQHHFKSVEA
                     SLWGKHLWSPSYFAGSCGDAPLEMIKQYIQDQETPH"
     gene            complement(3523. .4083)
                     /gene="HP1009"
     CDS             complement(3523. .4083)
                     /gene="HP1009"
                     /note="similar to GB:U03554 PID:520404 percent identity:
                     23.49; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="site-specific recombinase"
                     /protein_id="AAD08059.1"
                     /db_xref="GI:2314158"
                     /translation="MYCFYSLTISASHLKVKRLRLLNEEMLVNFLFELAKQRKINSMA
                     KYVMYIRQFFDYLDRTKHYEFYPSLKNIAFAKHKDNLPKHLNSKDLKSFIYTLINYRT
                     RSSYEKRNKCILLLIILGGLRKSEVFNLELRNIVLEKEHYILLIKGKNNKERKAFIKI
                     AQTDIDTLAPLIRILLESIAKNLLSH"
     tRNA            5273. .5347
```



```
                     /gene="tRNA-Pro-1"
                     /product="Pro GGG"
     gene            5273. .5347
                     /gene="tRNA-Pro-1"
     gene            5444. .7471
                     /gene="HP1010"
     CDS             5444. .7471
                     /gene="HP1010"
                     /note="similar to PID:755167 percent identity: 38.47;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="polyphosphate kinase (ppk)"
                     /protein_id="AAD08054.1"
                     /db_xref="GI:2314153"
                     /translation="MNRPFNRELSWLAFNTRVLNEAKDESLPLLERLKPLAIYDTNLD
                     EFYMIRVAGLKQLYEHKIASKGIDGASPEEQLEKIKHYLAHEIEERELEPQKIQALLF
                     KKGLCITPYNELNLEQKAKAKTYFKEQLYALVLPFKLDSSHTFPPLANLTPALPARIK
                     DKETQIISYALIKLPSFIFRFVELEKGLFVLAEEIVEAHLEELFLEHEILDCMAFRVT
                     CDADIAITEDEAHDYADLMSKSLRKRNQGEIVRLQTQKGSQELLKTLLASLRSFQTHS
                     YKKHKLTGMHIYKSAIMLNLGDLWELVNHSDFKALKSPNFTPKIHPHFNENDLFKSIE
                     KQDLLLFHPYESFEPVIDLIEQAASDPATLSIKMTLYRVGKHSPIVKALIEAASKIQV
                     SVLVELKARFDEESNLHWAKALERAGALVVYGVFKLKVHAKNLLITKKTDNQLRRFTH
                     LSTGNYNPLSAKVYTDVSFFSAKNEIANDIIKLFHSLLTSSATHSALFTLFMAPKQIK
                     PKIIELIQNEMNHQQEGYIILKANALVDSEIIEWLYQASQKGVKIDLIIRGICCLKPQ
                     VKGLSENIRVYSIVGKYLEHARIYYFKHENIYFSSADLMPRNLERRVELLIPATNPKI
                     AHKLLHILEIQLKDTLKRYELNSKGRYIKVSNPNDPLNSQDYFEKQALKTF"
     gene            7508. .8563
                     /gene="HP1011"
     CDS             7508. .8563
                     /gene="HP1011"
                     /note="similar to GB:L42023 SP:P45477 PID:1007463
                     PID:1221542 PID:1205636 percent identity: 41.52;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydroorotate dehydrogenase (pyrD)"
                     /protein_id="AAD08055.1"
                     /db_xref="GI:2314154"
                     /translation="MLYSLVKKYLFSLDAEDAHEKVCKILRMLSSSPFLCNLIDSQWG
                     YQNPKLENEILGLHFPNPLGLAAGFDKNASMLRALMAFGFGYLEAGTLTNEAQVGNER
                     PRLFRHIEEESLQNAMGFNNYGAILGVRSFKRFAPYKTPIGINLGKNKHIEQAHALED
                     YKAVLSKCLNIGDYITFNLSSPNTPNLRDLQNKAFVHELFCMAKEMTHKPLFLKIAPD
                     LETDDMLEIVNSAIGAGAHGIIATNTTIDKSLVFAPKEMGGLSGKCLTKKSREIFKEL
                     AKAFFNKSVLVSVGGISDAKEAYERIKMGASLLQIYSAFIYNGPNLCQNILKDLVKLL
                     QKDGFLSVKEAIGADLR"
     gene            8560. .9894
                     /gene="HP1012"
     CDS             8560. .9894
                     /gene="HP1012"
                     /note="similar to PID:1161059 percent identity: 29.64;
```



```
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="protease (pqqE)"
                     /protein_id="AAD08056.1"
                     /db_xref="GI:2314155"
                     /translation="MKHPSVKRLLGLSSVLLVTLGASMHAQSYLPKHESVTLKNGLQV
                     VSVPLENKTGVIEVDVLYKVGSRNETMGKSGIAHMLEHLNFKSTKNLKAGEFDKIVKR
                     FGGVSNASTSFDITRYFIKTSQANLDKSLELFAETMGSLNLKEDEFLPERQVVAEERR
                     WRTDNSPIGMLYFRFFNTAYVYHPYHWTPIGFMDDIQNWTLKDIKKFHSLYYQPKNAI
                     VLVVGDVNSQKVFELSKKHFESLKNLDEKAIPTPYMKEPKQDGARTAVVHKDGVHLEW
                     VALGYKVPAFKHKDQVALDALSRLLGEGKSSWLQSELVDKKRLASQAFSHNMQLQDES
                     VFLPIAGGNPNVKAEALQKEIVALLEKLKKGEITQAELDKLKINQKADFISNLESSSD
                     VAGLFADYLVQNDIQGLTDYQRQFLDLKVSDLVRVANEYFKDTQSTTVFLKP"
     gene            9909. .10811
                     /gene="HP1013"
     CDS             9909. .10811
                     /gene="HP1013"
                     /note="similar to GB:M12844 SP:P05640 PID:145708 GB:U00096
                     PID:1788823 percent identity: 39.51; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydrodipicolinate synthetase (dapA)"
                     /protein_id="AAD08057.1"
                     /db_xref="GI:2314156"
                     /translation="MQFHSSSALITPFKKDLSVDEAAYETLIKRQIFQGMDACVPVGT
                     TGESATLTHKEHMRCIEIAIETCKNTKTPSNSRMKVLAGVGSNATSESLSLAKFAQKI
                     GADAILCVSPYYNRPTQQGLFEHYKTIAQSVEIPVMLYDVPSRTGVSIEVPTALKLFR
                     EVPNIKAIKEASGSLKRVTELHYYEKDFKIFSGEDSLNHSIMPSGGCGVISVTGNLMP
                     NLISQMVNCALKQKYQQALEIQNKLFCLHQALFVETNPIPIKMAMHLAGLIENPSYRL
                     PLVAPSKETIQLLEKTLQQYEVIA"
     gene            10808. .11596
                     /gene="HP1014"
     CDS             10808. .11596
                     /gene="HP1014"

   Query Match           1.2%;  Score 73;  DB 1;  Length 11680;
   Best Local Similarity   100.0%;  Pred. No. 2.1e-23;
   Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4332 acgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaaaagacttaacc 4391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1622 ACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAAAAGACTTAACC 1681

Qy   4392 ctttgcttaaaat 4404
           |||||||||||||
Db   1682 CTTTGCTTAAAAT 1694

RESULT 159
```



```
AE000557
LOCUS       AE000557    16693 bp    DNA             BCT       06-APR-1999
DEFINITION  Helicobacter pylori 26695 section 35 of 134 of the complete genome.
ACCESSION   AE000557 AE000511
VERSION     AE000557.1  GI:2313514
KEYWORDS    .
SOURCE      Helicobacter pylori 26695.
  ORGANISM  Helicobacter pylori 26695
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 16693)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and
            Venter,J.C.
  TITLE     The complete genome sequence of the gastric pathogen Helicobacter
            pylori
  JOURNAL   Nature 388 (6642), 539-547 (1997)
  MEDLINE   97394467
  REMARK    Erratum:[[published erratum appears in Nature 1997 Sep
            25;389(6649):412]]
REFERENCE   2  (bases 1 to 16693)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C.,
            Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M.,
            Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 16693)
  AUTHORS   White,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .16693
                     /organism="Helicobacter pylori 26695"
                     /strain="26695"
                     /db_xref="taxon:85962"
     gene            113. .601
                     /gene="HP0408"
     CDS             113. .601
```

```
                     /gene="HP0408"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0408"
                     /protein_id="AAD07487.1"
                     /db_xref="GI:2313525"
                     /translation="MNIFQTSLKCCVGLVLSVGVLLGDSKAFKVRVDKSLTPPPLNVL
                     SLAFKQDMKKEVIFVITKSNKLSKKVLCDFDAFLLPETLMSGMPKKALFHKEFLFQSK
                     ENKTLYAFSLIDSQYCSKGGNYRYELEKLERWFVQKAPELAESYRVNYKNQYNKTQIS
                     QK"
     gene            612. .2138
                     /gene="HP0409"
     CDS             612. .2138
                     /gene="HP0409"
                     /note="similar to GP:460064 percent identity: 56.14;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="GMP synthase (guaA)"
                     /protein_id="AAD07477.1"
                     /db_xref="GI:2313515"
                     /translation="MILVLDFGSQYTQLIARRLRERGIYTEIVPFFESIENIQKKAPK
                     GLILSGGPASVYAKDAYKPSGKIFDLNVPILGICYGMQYLVDPFGGVVVGANEQEFGK
                     AVLEITQNSVIFEGVKIKSLVWMSHMDKVIELPKGFTTLAKSPNSPHCAIENGKIFGL
                     QFHPEVVQSEEGGKILENFALLVCGCEKTWGMQHFAQREIARLKEKIANAKVLCAVSG
                     GVDSTVVATLLHRAIKDNLIAVFVDHGLLRKNEKERVQAMFKDLKIPLNTIDAKEVFL
                     SKLKGVSEPELKRKIIGETFIEVFEKEAKKHHLKGKIEFLAQGTLYPDVIESVSVKGP
                     SKVIKTHHNVGGLPEWMDFKLIEPLRELFKDEVRLLGKELGVSQDFLMRHPFPGPGLA
                     VRILGEISESKIKRLQEADFIFIEELKKANLYDKVWQAFCVLLNVNSVGVMGDNRTYE
                     NAICLRAVNASDGMTASFSFLEHSFLEKVSNRITNEVSGINRVVYDITSKPPGTIEWE
                     "
     gene            2223. .2972
                     /gene="HP0410"
     CDS             2223. .2972
                     /gene="HP0410"
                     /note="similar to SP:Q48254 percent identity: 24.19;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="putative neuraminyllactose-binding hemagglutinin
                     homolog (hpaA)"
                     /protein_id="AAD07478.1"
                     /db_xref="GI:2313516"
                     /translation="MKKGSLAIVLGSLLASGAFYTALADGMPAKQQHNNTGESVELHF
                     HYPIKGKQEPKNSHLVVLIEPKIEINKVIPESYQKEPEKSLPLQLSSFLERKGYSVSQ
                     FKDASEIPQDIKEKALLVLRMDGNVAILEDIVEESDALSEEKVIDMSSGYLNLNFVEP
                     KSEDIIHSFGIDVSKIKAVIERVELRRTNSGGFVPKTFVHRIKETDHDQAIRKIMNQA
                     YHKVMVHITKELSKKHMEHYEKVSSEMKKRK"
     gene            3068. .3385
```



```
                     /gene="HP0411"
     CDS             3068. .3385
                     /gene="HP0411"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0411"
                     /protein_id="AAD07488.1"
                     /db_xref="GI:2313526"
                     /translation="MPLLILAVFLKNYAQKLENSAIITLHVKISGRSAVWLAHLVWDQ
                     GVEGSNPFAPTTFLHNLLFCLRFPGRISLPFRMVGVAQSVRASDCGPEGRGFDSHLPP
                     HLI"
     tRNA            3156. .3230
                     /gene="tRNA-Pro-2"
                     /product="Pro TGG"
     gene            3156. .3230
                     /gene="tRNA-Pro-2"
     tRNA            3299. .3372
                     /gene="tRNA-His-1"
                     /product="His GTG"
     gene            3299. .3372
                     /gene="tRNA-His-1"
     tRNA            3433. .3506
                     /gene="tRNA-Arg-2"
                     /product="Arg TCG"
     gene            3433. .3506
                     /gene="tRNA-Arg-2"
     tRNA            3532. .3605
                     /gene="tRNA-Arg-3"
                     /product="Arg TCT"
     gene            3532. .3605
                     /gene="tRNA-Arg-3"
     gene            3635. .3733
                     /gene="HP0412"
     CDS             3635. .3733
                     /gene="HP0412"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0412"
                     /protein_id="AAD07491.1"
                     /db_xref="GI:2313529"
                     /translation="MDSLEWGLLIFLILVFLVGIGYYIFMVFFSKD"
     gene            complement(4512. .5318)
                     /gene="HP0413"
     CDS             complement(4512. .5318)
                     /gene="HP0413"
                     /note="similar to PID:565287 percent identity: 32.46;
                     identified by sequence similarity; putative"
```



```
                     /codon_start=1
                     /transl_table=11
                     /product="transposase-like protein, PS3IS"
                     /protein_id="AAD07479.1"
                     /db_xref="GI:2313517"
                     /translation="MKVNKGPKFRLYPTKEQQDKLQHCFFVYNQAYNIGLNELQEQYE
                     TNKDSPPKERKYKKSSELDNAIKQCLRARDLPFSAVIAQQARMNVERALKDAFKVKNR
                     GFPKFKNSKSAKQSFSWNNQGFSIKESDDECFKTFTLMKMPLLMRMHRRLPPNFKVKQ
                     ISISCSHRKYFVSFSVEYEQDITPIKNTKNGVGLDLNILDTACSCEINNHDKLTDFKQ
                     YQTDMKELLGIEIDEELDTKRLIPTYSKLYSLKKYSKKFKRLQRKQSRRC"
     gene            5356. .5772
                     /gene="HP0414"
     CDS             5356. .5772
                     /gene="HP0414"
                     /note="similar to GB:L25848 GB:X52093 PID:1177217
                     PID:439619 PID:1150644 percent identity: 33.91; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="IS200 insertion sequence from SARA17"
                     /protein_id="AAD07480.1"
                     /db_xref="GI:2313518"
                     /translation="MKKIDDMRHGRHCVFLMHVHFVFVTKYRRSAFNKEVIDFLGSVF
                     ARVCKDPESELVEFDGESDHVHLLINYPPKVSVSKLVNSLKGVSSRLTRQHHFKSVEA
                     SLWGKHLWSPSVFAGSCGDAPLEMIKQYIQDQETPH"
     gene            6167. .8038
                     /gene="HP0415"
     CDS             6167. .8038
                     /gene="HP0415"
                     /note="similar to GP:1787591 percent identity: 44.44;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical integral membrane
                     protein"
                     /protein_id="AAD07481.1"
                     /db_xref="GI:2313519"
                     /translation="MRLLLWWVLVLSLFLNPLRAVEEHETDAVDLFLIFNQINQLNQV
                     IETYKKNPERSAEISLYNTQKNDLIKSLTSKVLNERDKIGIDINQNLKEQEKIKKRLS
                     KSINGDDFYTFMKDRLSLDILLIDEILYRFIDKIRSSIDIFSEQKDVESISDAFLLRL
                     GQFKLYTFPKNLGNVKMHELEQMFSDYELRLNTYTEVLRYIKNHPKEVLPKNLIMEVN
                     MDPVLNKISKVLPFTTHSLQVSKIVLALTILALLLGLRKLITHLLALLLDRIFEIMQR
                     NKKMHVNVQKSIVSPVSVFLALFSCDVALDIFYYPNASPPKVSMWVGAVYIWLLAWLV
                     IALFKGYGEALVTNMATKSTHNFRKEVINLILKVVYFLIPIVALLGVLKQLGFNVSAI
                     IASLGIGGLAVALAVKDVLANFFASVILLLDNSFSQGDWIVCGEVEGTVVEMGLRRTT
                     IRAFDNALLSVPNSELAGKPIRNWSRRKVGRRIKMEIGLTYSSSQSALQLCVKDIKEM
                     LENHPKIANGADSALQNVSDYRYMFKKDIVSIDDFLGYKNNLPVFLDQFADSSINILV
                     YCFSKTVVWEEWLEVKEDVMLKIMGIVEKHHLSFAFPSQSLYVESLPEVSLKEGAKI"

Query Match             1.2%;  Score 73;  DB 1;  Length 16693;
Best Local Similarity   100.0%;  Pred. No. 2e-23;
```



```
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    4332 acgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaaaagacttaacc 4391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3995 ACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAAAAGACTTAACC 4054

Qy    4392 ctttgcttaaaat 4404
           |||||||||||||
Db    4055 CTTTGCTTAAAAT 4067


RESULT 160
AF289440
LOCUS       AF289440      576 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR12 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289440
VERSION     AF289440.1  GI:9964234
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES                Location/Qualifiers
     source             1. .576
                        /organism="Helicobacter pylori"
                        /isolate="CR12"
                        /db_xref="taxon:210"
                        /country="Costa Rica"
                        /note="isolated from a patient with gastric
                        adenocarcinoma"
     gene               <1. .>576
                        /gene="cagA"
     CDS                <1. .>576
                        /gene="cagA"
                        /note="CagA"
                        /codon_start=1
                        /transl_table=11
```

```
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09889.1"
                     /db_xref="GI:9964235"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKDSTEPIYAKVNKRKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIAGGLGGVGQAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPFPLKRHDKVDDLSKVGLSREQQLKRKIDNLSQAVSEAKAGFFGNL
                     EQTIDKLKDSTKNNTVNLWAESAKKVPASFSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      207 a    100 c    129 g    140 t
ORIGIN


  Query Match             1.2%;  Score 70;  DB 2;  Length 576;
  Best Local Similarity   100.0%;  Pred. No. 6.7e-22;
  Matches   70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3255 tgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaa 3314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAA 200

Qy   3315 tcaaatagca 3324
          ||||||||||
Db    201 TCAAATAGCA 210


RESULT 161
AE001512
LOCUS       AE001512    10781 bp    DNA             BCT       20-JAN-1999
DEFINITION  Helicobacter pylori, strain J99 section 73 of 132 of the complete
            genome.
ACCESSION   AE001512 AE001439
VERSION     AE001512.1  GI:4155392
KEYWORDS    .
SOURCE      Helicobacter pylori J99.
  ORGANISM  Helicobacter pylori J99
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 10781)
  AUTHORS   Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C.,
            Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G.,
            Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C.,
            Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E.,
            Vovis,G.F. and Trust,T.J.
  TITLE     Genomic-sequence comparison of two unrelated isolates of the human
            gastric pathogen Helicobacter pylori
  JOURNAL   Nature 397 (6715), 176-180 (1999)
  MEDLINE   99120557
  REMARK    Erratum:[[published erratum appears in Nature 1999 Feb
            25:397(6721):719]]
```



```
REFERENCE   2  (bases 1 to 10781)
  AUTHORS   King,B.L., Alm,R.A. and Trust,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Address all correspondence to:  hp@arcb.us.astra.com  or  Richard
            A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
            MA,  02139.  Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,
            Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
            Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
            Corporation, 100 Beaver Street, Waltham, MA, 02453.  Qin Jiang and
            Diane E. Taylor are with the University of Alberta Department of
            Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G
            2H7 and the Canadian Bacterial Diseases Network.  All other authors
            are with Astra Research Center Boston, 128 Sidney Street,
            Cambridge, MA, 02139.  Putative identifications, sequence
            alignments, and name and sequence search capabiblity are available
            at ARCB's World Wide Web site. (URL:
            http://www.astra-boston.com/hpylori).
FEATURES            Location/Qualifiers
     source         1. .10781
                    /organism="Helicobacter pylori J99"
                    /strain="J99"
                    /db_xref="taxon:85963"
     gene           122. .646
                    /gene="jhp0824"
     CDS            122. .646
                    /gene="jhp0824"
                    /note="similar to H. pylori 26695 gene HP0891"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06410.1"
                    /db_xref="GI:4155401"
                    /translation="MPQIQSSHSNHFDFTIDTADRTKLLMSYLVVPTTANFNNVMHGG
                    ELLNLLDKVAYVCSTRYCAKGTVTLSVDGVTFKYPIPVGNLLTFLASINYVGNTSCEV
                    GIKVLSEDIKTREITHTNSCYFTMVAVENGKPTPMPKYEPKTEVEIRRYEGALKRKEM
                    RTRGYLKSGKHEGV"
     gene           complement(721. .1023)
                    /gene="jhp0825"
     CDS            complement(721. .1023)
                    /gene="jhp0825"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06402.1"
                    /db_xref="GI:4155393"
                    /translation="MLEIELKKKPTKDLKKHILNQKIELEVFDLVVENLRNQIPLDKR
                    FKDHALSGTYKGCRERHIKPDVLLVYRVKGNVLTLVRLGSHSELFCKPPTPLITLK"
     gene           complement(1172. .2500)
                    /gene="tnpB"
```



```
                    /note="jhp0826"
     CDS            complement(1172. .2500)
                    /gene="tnpB"
                    /codon_start=1
                    /transl_table=11
                    /product="IS606 TRANSPOSASE"
                    /protein_id="AAD06403.1"
                    /db_xref="GI:4155394"
                    /translation="MKVNKGFKFRLYPTKEQQDKLQHCFFVYNQAYNIGLNLLQEQYE
                    KNKDLPPKERTRKKSSELDKAIKHHLNARGLSFSSVIAQQSRMNVERALKDAFKVKNR
                    GFPKFKNSKSAKQSFSWNNQGFFIKESDEERFKIFTLMKMPLMMCMHRDPPPHSKVKQ
                    IVISCSHNRKYFVSFSVEYEQDITPIKNPKNGVGLDLNILDIACSCGVNNHKKLTDFKR
                    YSTDMKELLGIEIDEELDTKRLIPTYSKLYSLKKHSKKFRKLQRKQSRRVLKSKQNKT
                    KLGGNFYKTQKKLNQVFDKSSHQKTDRYHKITSELSKQFELIVVEDLQVKNMTKRAKL
                    KNVKQKSGLNQSILNTSFYQIISPLDYKQQHNGKLLVKVPPQYTSKTCHCCGNINHKL
                    KLNHRQYWCLECGYREHRDINAANNILSKGLSLFGVGNIHADFKEQSLSC"
     gene           2528. .2722
                    /gene="tnpA"
                    /note="jhp0827"
     CDS            2528. .2722
                    /gene="tnpA"
                    /codon_start=1
                    /transl_table=11
                    /product="IS606 TRANSPOSASE"
                    /protein_id="AAD06404.1"
                    /db_xref="GI:4155395"
                    /translation="MSVSKLVNSLKGVSSRLTRQHHFKSVEASLWGKHLWSPSYFAGS
                    CGGTPLEMIKQYIQEQETPH"
     gene           complement(2844. .3050)
                    /gene="jhp0828"
     CDS            complement(2844. .3050)
                    /gene="jhp0828"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06405.1"
                    /db_xref="GI:4155396"
                    /translation="MLKNVEMGANFYRELGKLEKQLAKYQSKVLEIKTQMKEIKKQYS
                    QAKKDENLANKYIPNDELKKDLSH"
     gene           complement(3118. .3264)
                    /gene="jhp0829"
     CDS            complement(3118. .3264)
                    /gene="jhp0829"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06406.1"
                    /db_xref="GI:4155397"
                    /translation="MKRHVIAFGLKIEILKKYKRTLQAHDDLRQLEPLGFENTQGSVY
                    LKDQ"
     gene           complement(3628. .3879)
```



```
                    /gene="jhp0830"
     CDS            complement(3628. .3879)
                    /gene="jhp0830"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06407.1"
                    /db_xref="GI:4155398"
                    /translation="MPNTTNKDYTKYSQKQLFNLINQLEQKISQAFDDKRGCCLGHEI
                    PNLETQQAIRGALNGENLEVIEDFSAWANEIKKEVNAEN"
     gene           complement(3927. .4199)
                    /gene="jhp0831"
     CDS            complement(3927. .4199)
                    /gene="jhp0831"
                    /note="similar to H. pylori 26695 gene HP0892"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06411.1"
                    /db_xref="GI:4155402"
                    /translation="MLTIETSKKFDKDLKILVKNGFDLKLLYKVVGNLATEQPLEPKY
                    KDHPLKGALKDFRECHLKPDLLLVYQIKKQENTLFLVRLGSHSELF"
     gene           complement(4213. .4500)
                    /gene="jhp0832"
     CDS            complement(4213. .4500)
                    /gene="jhp0832"
                    /note="similar to H. pylori 26695 gene HP0893"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06412.1"
                    /db_xref="GI:4155403"
                    /translation="MPNTTNKDYTKYSQKQLFNFLNSIKAKQKRALEKLKEIQTQKQR
                    IKKALQFKALHLTENGYTIEEERELARAKDTKNRLCFKSIEDFKKHCENL"
     gene           complement(4996. .7230)
                    /gene="babA"
                    /note="jhp0833"
     CDS            complement(4996. .7230)
                    /gene="babA"
                    /note="similar to H. pylori 26695 gene HP1243"
                    /codon_start=1
                    /transl_table=11
                    /product="outer membrane protein - adhesin"
                    /protein_id="AAD06409.1"
                    /db_xref="GI:4155400"
                    /translation="MKKHILSLTLGSLLVSTLSAEDDGFYTSVGYQIGEAAQMVTNTK
                    GIQDLSDRYESLNNLLNRYSTLNTLIKLSADPSAINAVRENLGASAKNLIGDKANSPA
                    YQAVLLAINAAVGFWNVVGYVTQCGGNANGQKSISSKTIFNNEPGYRSTSITCSLNGH
                    SPGYYGPMSIENFKKLNEAYQILQTALKRGLPALKENNGKVNVTYTYTCSGDGNNNCS
                    SQVTGVNNQKDGTKTKIQTIDGKSVTTTISSKVVDSRADGNTTGVSYTEITNKLEGVP
                    DSAQALLAQASTLINTINNACPYFHASNSSEANAPKFSTTTGKICGAPSEEISAIQKM
```

```
                     ITDAQELVNQTSVINEHEQTTPVGNNNGKPFNPFTDASFAQGMLANASAQAKMLNLAE
                     QVGQAINPERLSGTPQNPVKGFLATCNNPSTAGTGGTQGSAPGTVTTQTFASGCAYVG
                     QTITNLKNSIAHFCTQEQQIQQAENIADTLVNFKSRYSELGNTYNSITTALSNIPNAQ
                     SLQNAVSKKNNPYSPQGIDTNYYLNONSYMQIQTINQELGRNPFRKVGIVSSQTNNGA
                     MNGIGIQVGYKQPFGQKRKWGARYYGFFDYNHAFIKSSFFNSASDVWTYGFGADALYN
                     FINDKATNFLGKNNKLSVGLFGGIALAGTSWLNSEYVNLATMNNVYNAKMNVANFQFL
                     FNMGVRMNLARPKKKDSDHAAQHGIELGLKIPTINTNYYSFMGAELKYRRLYSVYLNY
                     VFAY"
     gene            complement(8227. .8637)
                     /gene="jhp0834"
     CDS             complement(8227. .8637)
                     /gene="jhp0834"
                     /codon_start=1
                     /transl_table=11
                     /product="putative"
                     /protein_id="AAD06408.1"
                     /db_xref="GI:4155399"
                     /translation="MCGVLSLQARLSFENCSKPLDCSLFATTCTPQNPIGSCMVSSQG
                     GVCGVLSLQARLSFENCSKPLDCSLFATTCTPQNPIGSCMVSSQGGVRGVLSLQARLS
                     FKNCPLVKSLKPSQLTLMLTGFLGGFSPLTTGFW"
     gene            complement(8604. .9716)
                     /gene="hypD"
                     /note="jhp0835"
     CDS             complement(8604. .9716)
                     /gene="hypD"

  Query Match             1.2%;  Score 70;  DB 1;  Length 10781;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-22;
  Matches   70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4332 acgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaaaagacttaacc 4391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1177 ACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAAAAGACTTAACC 1236

Qy   4392 ctttgcttaa 4401
          ||||||||||
Db   1237 CTTTGCTTAA 1246


RESULT 162
AF198473
LOCUS       AF198473      291 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain Peru24C cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198473
VERSION     AF198473.1  GI:7248672
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
```



```
REFERENCE   1  (bases 1 to 291)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulat,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 291)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /strain="Peru24C"
                     /db_xref="taxon:210"
                     /country="Peru"
                     /note="European type"
     gene            <1. .>291
                     /gene="cagA"
     CDS             <1. .>291
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43613.1"
                     /db_xref="GI:7248673"
                     /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSEK"
BASE COUNT      106 a      51 c      68 g      66 t
ORIGIN


  Query Match             1.2%;  Score 69;  DB 1;  Length 291;
```



```
  Best Local Similarity   100.0%;  Pred. No. 2.1e-21;
  Matches   69;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1716 aaacaaaatagatttcatggaatttcttgcacaaaataatgctaaattagacaacttgag 1775
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    223 AAACAAAATAGATTTCATGGAATTTCTTGCACAAAATAATGCTAAATTAGACAACTTGAG 282

Qy   1776 cgagaaaga 1784
          |||||||||
Db    283 CGAGAAAGA 291


RESULT 163
HPU95957/c
LOCUS       HPU95957      2378 bp    DNA             BCT       01-FEB-1999
DEFINITION  Helicobacter pylori insertion sequence IS606 transposase homologs A
            (tnpA) and B (tnpB) genes, complete cds.
ACCESSION   U95957
VERSION     U95957.1  GI:2114469
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 2378)
  AUTHORS   Kersulyte,D., Akopyants,N.S., Clifton,S.W., Roe,B.A. and Berg,D.E.
  TITLE     Novel sequence organization and insertion specificity of IS605 and
            IS606: chimaeric transposable elements of Helicobacter pylori
  JOURNAL   Gene 223 (1-2), 175-186 (1998)
  MEDLINE   99077840
REFERENCE   2  (bases 1 to 2378)
  AUTHORS   Akopyants,N.S. and Berg D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-1997) Molecular Microbiology, Washington
            University School of Medicine, 660 South Euclid Ave., St.Louis, MO
            63110, USA
FEATURES             Location/Qualifiers
     source          1. .2378
                     /organism="Helicobacter pylori"
                     /strain="84-183"
                     /db_xref="taxon:210"
                     /map="near right end of the cag pathogenicity island in
                     strain 84-183; approximately 0.7 kb downstream of 3' end
                     of cagA gene"
     source          286. .2253
                     /organism="Helicobacter pylori"
                     /strain="84-183"
                     /db_xref="taxon:210"
                     /insertion_seq="IS606"
                     /note="similar to H. pylori transposable element IS605,
                     GenBank Accession Numbers AC000108 and U60177"
```



```
     gene            complement(417. .833)
                     /gene="tnpA"
     CDS             complement(417. .833)
                     /gene="tnpA"
                     /note="TnpA; IS200 transposase homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase homolog A"
                     /protein_id="AAD11513.1"
                     /db_xref="GI:2114470"
                     /translation="MKKIDDMRHGRHCVFLMHTHLVFVTKYRRKAFNKEVIDFLGSVF
                     AKVCKDFESELVEFDGESDHVHLLINYPPKVSVSKLVNSLKGVSSRLTRQHHFKSVEA
                     SLWGKHLWSPSYFAGSCGGAPLEMIKQYIQEQETPH"
     gene            870. .2198
                     /gene="tnpB"
     CDS             870. .2198
                     /gene="tnpB"
                     /note="TnpB; IS1341 transposase homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase homolog B"
                     /protein_id="AAD11514.1"
                     /db_xref="GI:2114471"
                     /translation="MKVNKGFKFRLYPTKEQQDKLQRCFFVYNQAYNIGLNLLQEQYE
                     TNKDSPPKERKWKKSSELDKAIKHHLNARGLGFSGVIAQQSRMNVERALKDAPKVKDR
                     GFPKFFNSKSAKQSFSWNNQGFSIKDSDEERFKIPTLMKMPLMMRMHRDFPPHSKVKQ
                     IVISWSHRKYFVSFCVEYEQDITPIKNPKNGVGLDLNILDIACSCGVNNHKKLTDFKQ
                     YPTDMKELLGIEIDEELDTKRLIPTYSKLYSLKKYSKKFKRLQRKQSRRVLKSKQNKT
                     KLGGNFYKTQKKLNQAFDKSSHQKTDRYHKITSELSKQFELVVVEDLQVKNMTKRAKL
                     KNVKQKSGLNQSILNTSFYQIISFLDYKQQHNGKLLVKVPPQYTSKTCHCCGNINHKL
                     KLNHRQYWCLECGYREHRDINAANNIISKGLSLFGVGNIHADFKEQSLSC"
BASE COUNT      822 a    456 c    399 g    701 t
ORIGIN


  Query Match             1.2%;  Score 69;  DB 2;  Length 2378;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-21;
  Matches   69;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4332 acgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaaaagacttaacc 4391
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2193 ACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAAAAGACTTAACC 2134

Qy   4392 ctttgctta 4400
          |||||||||
Db   2133 CTTTGCTTA 2125


RESULT 164
AF190660/c
LOCUS       AF190660      2963 bp    DNA             BCT       04-OCT-2000
```

```
DEFINITION  Helicobacter pylori strain 84-183 cag pathogenicity island
            polymorphic right end, type Ia motif; insertion sequence IS606,
            complete sequence; and glutamate racemase (glr) gene, partial cds.
ACCESSION   AF190660
VERSION     AF190660.1  GI:10567547
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 2963)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
  PUBMED    10809702
REFERENCE   2  (bases 1 to 2963)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .2963
                     /organism="Helicobacter pylori"
                     /strain="84-183"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="USA"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ic motif; similar to
                     Helicobacter pylori strain Gambia94/24 type Ia motif
                     presented in GenBank Accession Number AF190658"
     misc_feature    26. .157
                     /note="similar to Helicobacter pylori insertion sequence
                     IS606"
     gene            210. .640
                     /gene="putative DNA helicase"
                     /note="truncated; similar to Helicobacter pylori strain
                     26695 putative helicase gene HP0548"
     source          739. .2705
                     /organism="Helicobacter pylori"
                     /strain="84-183"
                     /db_xref="taxon:210"
```



```
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /insertion_seq="IS606"
                     /country="USA"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type Ic motif; similar to type Ia
                     motif; similar to Helicobacter pylori insertion sequence
                     IS606 presented in GenBank Accession Number U95957"
     repeat_region   2765. .2963
                     /note="defines the terminus of the cag pathogenicity
                     island"
                     /rpt_type=direct
     gene            complement(2765. .>2963)
                     /gene="glr"
                     /note="flanks cag pathogenicity island"
     CDS             complement(2765. .>2963)
                     /gene="glr"
                     /note="similar to Helicobacter pylori strain 26695
                     glutamate racemase HP0549"
                     /codon_start=2
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG18483.1"
                     /db_xref="GI:10567548"
                     /translation="IEGYPMGHFALPTPPLLIHSGDAIVEYLQQKYALKKNAHAFPKV
                     EFHASGDVIWLEKQAKEWLKL"
BASE COUNT     1027 a     566 c     521 g     849 t
ORIGIN


  Query Match             1.2%;  Score 69;  DB 1;  Length 2963;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-21;
  Matches   69;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    4332 acgaaaggcttttgttctttaaagtctgcatggatatttcctaccccaaaaagacttaacc 4391
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2643 ACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAAAAGACTTAACC 2584

Qy    4392 ctttgctta 4400
           |||||||||
Db    2583 CTTTGCTTA 2575


RESULT 165
AF084492
LOCUS       AF084492      419 bp    DNA             BCT       05-MAR-1999
DEFINITION  Helicobacter pylori strain Lith4-1 glutamate racemase gene, partial
            cds.
ACCESSION   AF084492
VERSION     AF084492.1  GI:3719310
KEYWORDS    .
SOURCE      Helicobacter pylori.
```



```
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Kersulyte,D., Chalkauskas,H. and Berg,D.E.
  TITLE     Emergence of recombinant strains of Helicobacter pylori during
            human infection
  JOURNAL   Mol. Microbiol. 31 (1), 31-43 (1999)
  MEDLINE   99141580
REFERENCE   2  (bases 1 to 419)
  AUTHORS   Kersulyte,D., Chalkauskas,H. and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-1998) Molecular Microbiology, Washington
            University School of Medicine, 4566 Scott Avenue, St. Louis, MO
            63110, USA
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Helicobacter pylori"
                     /strain="Lith4-1"
                     /db_xref="taxon:210"
                     /note="cag- strain"
     misc_feature    1. .419
                     /note="motif a; flanking region of cag pathogenicity
                     island insertion site in cag+ strains"
     CDS             complement(224. .>419)
                     /note="similar to Helicobacter pylori strain 26695 HP0549"
                     /codon_start=2
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAC63302.1"
                     /db_xref="GI:3719311"
                     /translation="EGYFMGHFAXPTPPLLIHSGDAIVEYLQQKYALKKNAHAFPKVE
                     FHASGDVIWLEKQAKEWLKL"
BASE COUNT      132 a      90 c      75 g     121 t       1 others
ORIGIN


  Query Match             1.1%;  Score 68;  DB 1;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-21;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5075 tgtttttctagccagatcacatcgccgctcgcatgaaattccactttagggaatgcgtgt 5134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 TGTTTTTCTAGCCAGATCACATCGCCGCTCGCATGAAATTCCACTTTAGGGAATGCGTGT 308

Qy    5135 gcatttttt 5142
           |||||||||
Db     309 GCATTTTTT 316


RESULT 166
```



```
AF201074
LOCUS       AF201074      430 bp    DNA             BCT       14-SEP-2000
DEFINITION  Helicobacter pylori strain India48A cag pathogenicity island
            polymorphic right end, type V motif and glutamate racemase (glr)
            gene, partial cds.
ACCESSION   AF201074
VERSION     AF201074.1  GI:10121331
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
  PUBMED    10809702
REFERENCE   2  (bases 1 to 430)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Helicobacter pylori"
                     /strain="India48A"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="India: Calcutta"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; isolated from gastric ulcer patient"
     misc_feature    1. .197
                     /note="type V motif"
     misc_feature    29. .149
                     /note="similar to Helicobacter pylori insertion sequence
                     IS606 presented in GenBank Accession Number U95957"
     variation       150. .191
                     /note="absent in type I, II, III and IV motifs"
                     /replace=""
     repeat_region   200. .430
                     /note="defines the terminus of the cag pathogenicity
                     island"
```

```
                     /rpt_type=direct
     gene            complement(200. .>430)
                     /gene="glr"
                     /note="flanks cag pathogenicity island"
     CDS             complement(200. .>430)
                     /gene="glr"
                     /note="similar to Helicobacter pylori strain 26695
                     glutamate racemase HP0549"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG13132.1"
                     /db_xref="GI:10121332"
                     /translation="GCTHFPLIAQKIEGYFMEHFALSTPPLLIHSGDAIVEYLQQKYA
                     LKKNAHAFPKVEPHASGDVIWLEKQAKEWLKL"
BASE COUNT      132 a     84 c     79 g    135 t
ORIGIN


  Query Match             1.1%;  Score 68;  DB 1;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-21;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5075 tgtttttctagccagatcacatcgccgctcgcatgaaattccactttagggaatgcgtgt 5134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 TGTTTTTCTAGCCAGATCACATCGCCGCTCGCATGAAATTCCACTTTAGGGAATGCGTGT 284

Qy   5135 gcatttttt 5142
           |||||||||
Db    285 GCATTTTT 292


RESULT 167
AF191010
LOCUS       AF191010      589 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain NCTC 11637 cag pathogenicity island
            polymorphic right end, type IIIa motif.
ACCESSION   AF191010
VERSION     AF191010.1  GI:10567573
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 589)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
```



```
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 589)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .589
                     /organism="Helicobacter pylori"
                     /strain="NCTC 11637"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="Australia"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type IIIa motif; similar to
                     Helicobacter pylori strain Sweden53 type IIIa motif
                     presented in GenBank Accession Number AF190662; 39 bp
                     deletion (absent in strain Sweden53) located between the
                     remnant of IS606 and function unknown sequence; 29 bp
                     deletion follows first one after 9 bp"
BASE COUNT      186 a     98 c    119 g    186 t
ORIGIN


  Query Match             1.1%;  Score 68;  DB 1;  Length 589;
  Best Local Similarity   100.0%;  Pred. No. 6e-21;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5075 tgtttttctagccagatcacatcgccgctcgcatgaaattccactttagggaatgcgtgt 5134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427 TGTTTTTCTAGCCAGATCACATCGCCGCTCGCATGAAATTCCACTTTAGGGAATGCGTGT 486

Qy   5135 gcatttttt 5142
           |||||||||
Db    487 GCATTTTT 494


RESULT 168
AF249275
LOCUS       AF249275     3524 bp    DNA             BCT       12-JUL-2000
DEFINITION  Helicobacter pylori strain CAPM N62 cytotoxin associated protein
            (cagA) gene, complete cds.
ACCESSION   AF249275
VERSION     AF249275.1  GI:9049777
KEYWORDS    .
```



```
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 3524)
  AUTHORS   Zhou,J.-C. and Zhang,J.-Z.
  TITLE     Sequencing and analysis of cagA genes of Helicobacter pylori
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3524)
  AUTHORS   Zhang,J.-Z. and Zhou,J.-C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2000) Dept. of Diarrhea Disease, Institute of
            Epidemiology and Microbiology, Chinese Academy of Preventive
            Medicine, P.O. Box 5, Changping, Beijing 102206, China
FEATURES             Location/Qualifiers
     source          1. .3524
                     /organism="Helicobacter pylori"
                     /strain="CAPM N62"
                     /db_xref="taxon:210"
     gene            1. .3522
                     /gene="cagA"
     CDS             1. .3522
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein"
                     /protein_id="AAF82477.1"
                     /db_xref="GI:9049778"
                     /translation="MTNETIDQTTTPDQTLNQTDPVPQRFINNLQVAFIKVDNAVASF
                     DPDQKPIVDKNDRDNRQAFEKISQLREEYANKAIKNPAKKNQYFSDFINKSNDLINKD
                     NLIAVDSSVESFKKFGDQRYQIFTSWVSLQKDPSEINTQQIRDFMENIIQPPISDDKE
                     KAEFLRSAKQSFAGIIIGNQIRSDEKFMGVFDESLKARQEAEKNEEPAGGDWLDIFLS
                     FVFNKKQSSDLKETLNQEPRPDFEQNIATTTTDIQGLPPEARDLLDERGNPSKFTLGD
                     MEHLDVEGVADKDPNVKFNQLLIHNNALSSVLMGGHSNIEPEKVSLLYGDNGGPEARH
                     DWNATVGYKDQQGSNVATLINAHLNNGSGLVIAGNEDGIKNPSFYLYKEDQLTGLKQA
                     MSQEEIQNKVDFMEFLAKNNAKLDNLSEKEKEKFQTEIGNFQKDRKAYLDALGNDHVA
                     FVSKKDPKHLALVTEFGNGELSYTLKDYGKKQDRALDGETKTTLQGSLKYDGVMFVNY
                     SNFKYTNASKSPDKGVGATNGVSHLEANPSKVAVFNLPNLNNLAITNYIRRDLEDKLW
                     AKGLSSQEANKLIKDFLNSNKEMVGKVLNFNKAVAEAKNTGNYDGVKKAQKDLEKSLR
                     KREHLEKEVAKKLESRNDNKNRMEAKAQANSQKDKIFALINQEASKEARAAAFDPNLK
                     GVRSELSDKLENINKNLKDFGKSFDELKNGKNNDFGKAEETLKTLKDSVKDLGINPEW
                     ISKIENLNAALNDFKNGKNKDFSKVTQAKSDLENSIKDVIINQKITDKVDNLNQAVSE
                     TKLTGDFSKVEQALAELKNLSLDLGKNSDLQSVKNSVNGTLVGNGLSQAEATTLTKNF
                     SDIKKELNEKLFGNSNNNNGLKNEDIVAQVNKKKAGQTASPEEPIYAQVAKKVSAKI
                     DQLNEATSAINRKIDRINKIASAGKGVGGFSGAGRSVSPEPIYATIDFDDANQAGFPL
                     RRSAAVNGLSKVGLSREQELTRRIGDLNQAASEAKTGHFDKLEQKIDELKDSTKKNAL
                     KLWVENAKQVPTSLQAKLDNYATNSHTRINSNVQHGTINEKATGMLTQKNPEWLKLVN
                     DKIVAHNVGSTHLSEYDKIGFNQKNMKDYSDSPKFSTKLNNAVKDIKSSFVQFLTNTF
                     STGSYSLMEANVEHGVKNTTKSGFQKS"
BASE COUNT     1322 a    612 c    710 g    880 t
ORIGIN
```



```
  Query Match             1.1%;  Score 68;  DB 1;  Length 3524;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-21;
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2352 agtgaaaaaagctcagaaagatctttgaaaaatctctaaggaaacgagagcatttagagaa 2411
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1833 AGTGAAAAAAGCTCAGAAAGATCTTGAAAAATCTCTAAGGAAACGAGAGCATTTAGAGAA 1892

Qy   2412 agaagtag 2419
           ||||||||
Db   1893 AGAAGTAG 1900


RESULT 169
AF289444
LOCUS       AF289444      564 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR21 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289444
VERSION     AF289444.1  GI:9964242
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 564)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Helicobacter pylori"
                     /isolate="CR21"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>564
                     /gene="cagA"
     CDS             <1. .>564
```

```
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09893.1"
                     /db_xref="GI:9964243"
                     /translation="NFSDIKKELNAKLGNFNNNNNNGLKNEPIYAQVNKKKAGQVVSP
                     EEPIYAQVAKKVNAKIDQLNQAASGLGGVGQAGPPLKRHDKVDDLSKVGRSVSPEPIY
                     ATIDDLGGPFPLKKHAKVEDLSKVGLSREQELTQKIDNLNQAVSEAKAGFFGNLEQTI
                     DNLKDSTKHNPMNLWAESAKKVPASLSA"
BASE COUNT      202 a    103 c    125 g    134 t
ORIGIN


  Query Match           1.1%;  Score 67;  DB 2;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-20;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3428 ctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggttttttggca 3487
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    404 CTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTTTTGGCA 463

Qy   3488 atctaga 3494
           |||||||
Db    464 ATCTAGA 470


RESULT 170
AF289448
LOCUS       AF289448      572 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR35 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289448
VERSION     AF289448.1  GI:9964250
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patienta in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 572)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Helicobacter pylori"
                     /isolate="CR35"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>572
                     /gene="cagA"
     CDS             <1. .>572
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09897.1"
                     /db_xref="GI:9964251"
                     /translation="NFSDIKKELNAKLPGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGKAAGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPFPLKRHDKVDDLSKVGLSREQQLKQKIDNLSQAVSEAKAGFPGNL
                     EQTIDKLKDSTKNNTVNLWAESAKKVPASL"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      211 a     96 c    124 g    141 t
ORIGIN


  Query Match           1.1%;  Score 67;  DB 2;  Length 572;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-20;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 216

Qy   3331 ttgggtg 3337
           |||||||
Db    217 TTGGGTG 223


RESULT 171
AF289445
LOCUS       AF289445      576 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR22 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289445
VERSION     AF289445.1  GI:9964244
```



```
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 576)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Helicobacter pylori"
                     /isolate="CR22"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with gastric
                     adenocarcinoma"
     gene            <1. .>576
                     /gene="cagA"
     CDS             <1. .>576
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09894.1"
                     /db_xref="GI:9964245"
                     /translation="NFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQITGGLGGVRASTGFPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPFPLKRHDKVDDLSKVGLSREQQLKRKIDNLSQAVSEAKAGFPGNL
                     EQTIDKLKDSTKNNTVNLWAESAKKVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      210 a    100 c    126 g    140 t
ORIGIN


  Query Match           1.1%;  Score 67;  DB 2;  Length 576;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-20;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy   3255 tgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaa 3314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAA 200

Qy   3315 tcaaata 3321
           |||||||
Db    201 TCAAATA 207


RESULT 172
AF289459
LOCUS       AF289459      678 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR55 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289459
VERSION     AF289459.1  GI:9964272
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 678)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Helicobacter pylori"
                     /isolate="CR55"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>678
                     /gene="cagA"
     CDS             <1. .>678
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09908.1"
```

```
                     /db_xref="GI:9964273"
                     /translation="NPSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQAAGPPLKRHDKVDDLSKVGRSVSP
                     EPIYATIDDLGGPPFLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPPFLKRHDKVDD
                     LSKVGLSREKQLKQKIDNLSQAVSEAKAGFFGNLEQTIDKLKDSTKNNTVNLWAESAK
                     KVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      240 a    117 c    151 g    170 t
ORIGIN


  Query Match             1.1%;  Score 67;  DB 2;  Length 678;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-20;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3271 gctcaagttgctaaaaaggtaaatgcaaaaattgaccgactcaatcaaatagcaagtggt 3330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     157 GCTCAAGTTGCTAAAAAGGTAAATGCAAAAATTGACCGACTCAATCAAATAGCAAGTGGT 216

Qy    3331 ttgggtg 3337
           |||||||
Db     217 TTGGGTG 223


RESULT 173
HPY252965
LOCUS       HPY252965     244 bp    DNA             BCT       08-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
            protein A.
ACCESSION   AJ252965
VERSION     AJ252965.1  GI:5851961
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
REFERENCE   2  (bases 1 to 244)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
```



```
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Helicobacter pylori"
                     /strain="Dutch 161A"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .244
                     /gene="cagA"
     CDS             <1. .>244
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB55441.1"
                     /db_xref="GI:5851962"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGLKRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       94 a     38 c     53 g     59 t
ORIGIN


  Query Match             1.1%;  Score 65;  DB 2;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-19;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1716 aaacaaaatagatttcatggaatttcttgcacaaaataatgctaaattagacaacttgag 1775
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 AAACAAAATAGATTTCATGGAATTTCTTGCACAAAATAATGCTAAATTAGACAACTTGAG 239

Qy    1776 cgaga 1780
           |||||
Db     240 CGAGA 244


RESULT 174
HPY252966
LOCUS       HPY252966     244 bp    DNA             BCT       08-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
            protein A.
ACCESSION   AJ252966
VERSION     AJ252966.1  GI:5851963
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
```



```
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
REFERENCE   2  (bases 1 to 244)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Helicobacter pylori"
                     /strain="Dutch 161L"
                     /db_xref="taxon:210"
                     /country="Netherlands"
     gene            1. .244
                     /gene="cagA"
     CDS             <1. .>244
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB55442.1"
                     /db_xref="GI:5851964"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGLKRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       94 a     38 c     53 g     59 t
ORIGIN


  Query Match             1.1%;  Score 65;  DB 2;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-19;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1716 aaacaaaatagatttcatggaatttcttgcacaaaataatgctaaattagacaacttgag 1775
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 AAACAAAATAGATTTCATGGAATTTCTTGCACAAAATAATGCTAAATTAGACAACTTGAG 239

Qy    1776 cgaga 1780
           |||||
Db     240 CGAGA 244


RESULT 175
AF289464
LOCUS       AF289464      513 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR61 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289464
VERSION     AF289464.1  GI:9964282
KEYWORDS
SOURCE      Helicobacter pylori.
```



```
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 513)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Helicobacter pylori"
                     /isolate="CR61"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>513
                     /gene="cagA"
     CDS             <1. .>513
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09913.1"
                     /db_xref="GI:9964283"
                     /translation="NPSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     VSPEPIYATIDDLGGPPFLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPPFLKRHDK
                     VDDLSKVGLSREQQLKQKIDNLNQAVSEAKAGFFGNLEQTIDKLKDSTKNNTVNLWAE
                     SVKKVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      187 a     90 c    105 g    131 t
ORIGIN


  Query Match             1.1%;  Score 65;  DB 2;  Length 513;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-19;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3430 cagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggttttttggcaat 3489
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    355 CAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTTTTTGGCAAT 414

Qy   3490 ctaga 3494
          |||||
Db    415 CTAGA 419


RESULT 176
AF190662
LOCUS       AF190662      653 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain Sweden53 cag pathogenicity island
            polymorphic right end, type IIIa motif and glutamate racemase (glr)
            gene, partial cds.
ACCESSION   AF190662
VERSION     AF190662.1  GI:10567551
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A.,,Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 653)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .653
                     /organism="Helicobacter pylori"
                     /strain="Sweden53"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; type IIIa motif"
     misc_feature    11. .142
                     /note="similar to Helicobacter pylori insertion sequence
                     IS606 presented in GenBank Accession Number U95957"
     misc_feature    204. .453
```



```
                     /note="this sequence differs from type I strains such as
                     J99 and Gambia which contain a putative helicase gene at
                     this location"
     repeat_region   430. .653
                     /note="defines the terminus of the cag pathogenicity
                     island"
                     /rpt_type=direct
     gene            complement(460. .>653)
                     /gene="glr"
                     /note="flanks cag pathogenicity island"
     CDS             complement(460. .>653)
                     /gene="glr"
                     /note="similar to Helicobacter pylori strain 26695
                     glutamate racemase HP0549"
                     /codon_start=3
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG18485.1"
                     /db_xref="GI:10567552"
                     /translation="GYFMEHFALPTPPLLIHSGDAIVEYLQQKYALKKNAHAFPKVEF
                     HASGDVIWLEKQAKEWLKL"
BASE COUNT      189 a    107 c    149 g    208 t
ORIGIN


  Query Match             1.1%;  Score 65;  DB 1;  Length 653;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-19;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4308 cttaaaatatcccgacagacactaacgaaaggctttgttctttaaagtctgcatggatat 4367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     46 CTTAAAATATCCCGACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATAT 105

Qy   4368 ttcct 4372
          |||||
Db    106 TTCCT 110


RESULT 177
HPY239720
LOCUS       HPY239720     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain N6.
ACCESSION   AJ239720
VERSION     AJ239720.1  GI:4467758
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
```



```
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Helicobacter pylori"
                     /strain="N6"
                     /db_xref="taxon:210"
                     /country="France"
     gene            1. .243
                     /gene="cagA"
     CDS             <1. .>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37812.1"
                     /db_xref="GI:4467759"
                     /db_xref="SPTREMBL:Q9Z365"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       93 a     40 c     52 g     58 t
ORIGIN


  Query Match             1.1%;  Score 64;  DB 2;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-19;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1716 aaacaaaatagatttcatggaatttcttgcacaaaataatgctaaattagacaacttgag 1775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 AAACAAAATAGATTTCATGGAATTTCTTGCACAAAATAATGCTAAATTAGACAACTTGAG 239

Qy   1776 cgag 1779
          ||||
Db    240 CGAG 243


RESULT 178
HPY239731
LOCUS       HPY239731     243 bp    DNA             BCT       30-JUN-1999
DEFINITION  Helicobacter pylori partial cagA gene, strain SS1.
ACCESSION   AJ239731
```



```
VERSION     AJ239731.1  GI:4467780
KEYWORDS    cagA gene; cytotoxin associated protein A.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Achtman,M., Azuma,T., Berg,D.E., Ito,Y., Morelli,G., Pan,Z.J.,
            Suerbaum,S., Thompson,S.A., Van Der Ende,A. and Van Doorn,L.J.
  TITLE     Recombination and clonal groupings within helicobacter pylori from
            different geographical regions
  JOURNAL   Mol. Microbiol. 32 (3), 459-470 (1999)
  MEDLINE   99255683
REFERENCE   2  (bases 1 to 243)
  AUTHORS   van der Ende,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) van der Ende A., Medical Microbiology,
            Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Helicobacter pylori"
                     /strain="SS1"
                     /db_xref="taxon:210"
                     /country="Australia"
     gene            1. .243
                     /gene="cagA"
     CDS             <1. .>243
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="CAB37823.1"
                     /db_xref="GI:4467781"
                     /db_xref="SPTREMBL:Q9Z365"
                     /translation="VGYKDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKE
                     DQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT       93 a     40 c     52 g     58 t
ORIGIN


  Query Match             1.1%;  Score 64;  DB 2;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-19;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1716 aaacaaaatagatttcatggaatttcttgcacaaaataatgctaaattagacaacttgag 1775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 AAACAAAATAGATTTCATGGAATTTCTTGCACAAAATAATGCTAAATTAGACAACTTGAG 239

Qy   1776 cgag 1779
          ||||
```

```
Db    240 CGAG 243


RESULT 179
AF198478
LOCUS       AF198478      286 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain Peru8C cytotoxin associated protein CagA
            (cagA) gene, partial cds.
ACCESSION   AF198478
VERSION     AF198478.1  GI:7248682
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 286)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 286)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .286
                     /organism="Helicobacter pylori"
                     /strain="Peru8C"
                     /db_xref="taxon:210"
                     /country="Peru"
                     /note="European type"
     gene            <1. .>286
                     /gene="cagA"
```



```
     CDS             <1. .>286
                     /gene="cagA"
                     /note="similar to Helicobacter pylori strain 26695 HP0547"
                     /codon_start=2
                     /transl_table=11
                     /product="cytotoxin associated protein CagA"
                     /protein_id="AAF43618.1"
                     /db_xref="GI:7248683"
                     /translation="GNGGPGARHDWNATVGYKDQQGDNVATIINVHMKNGSGLVIAGG
                     EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKVDFMEFLAQNNAKLDNLSV"
BASE COUNT      100 a      54 c      68 g      64 t
ORIGIN


  Query Match             1.0%;  Score 62;  DB 1;  Length 286;
  Best Local Similarity   99.1%;  Pred. No. 4.5e-18;
  Matches  112;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1585 gtgcatatgaaaaacggcagtggcttagtcatagcaggtggtgagaaagggattaacaac 1644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     92 GTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCAGGTGGTGAGAAAGGGATTAATAAC 151

Qy   1645 ccctagtttttatctctacaaagaagaccaactcacaggctcacaacgagcatt 1697
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 CCTAGTTTTTATCTCTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATT 204


RESULT 180
HPY269876
LOCUS       HPY269876     449 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            US123 from United States.
ACCESSION   AJ269876
VERSION     AJ269876.1  GI:5921329
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 449)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
```



```
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Helicobacter pylori"
                     /isolate="US123"
                     /db_xref="taxon:210"
                     /country="USA"
     gene            1. .449
                     /gene="cagA"
     CDS             1. .>449
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56488.1"
                     /db_xref="GI:5921330"
                     /translation="MTNETINQQPQTEAAFNPQQFINNLQVAFIKVDNVVASFDPDQK
                     PIVDRNDRDNRQAFEKISQLREEFANKAIKNPAKKNQYFSSFISKSSDLINKDSLIDT
                     GSSIKSFQKFGTQRYQIFMNWVSHQKDPSKINTQKIRGFMENIIQPP"
BASE COUNT      166 a      88 c      73 g     122 t
ORIGIN


  Query Match             1.0%;  Score 61;  DB 2;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-17;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 ACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTTCAAG 79

Qy    614 t 614
          |
Db     80 T 80


RESULT 181
AE000537
LOCUS       AE000537    12901 bp    DNA             BCT       06-APR-1999
DEFINITION  Helicobacter pylori 26695 section 15 of 134 of the complete genome.
ACCESSION   AE000537 AE000511
VERSION     AE000537.1  GI:2313247
KEYWORDS    .
SOURCE      Helicobacter pylori 26695.
  ORGANISM  Helicobacter pylori 26695
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 12901)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
```



```
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and
            Venter,J.C.
  TITLE     The complete genome sequence of the gastric pathogen Helicobacter
            pylori
  JOURNAL   Nature 388 (6642), 539-547 (1997)
  MEDLINE   97394467
  REMARK    Erratum:[[published erratum appears in Nature 1997 Sep
            25;389(6649):412]]
REFERENCE   2  (bases 1 to 12901)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C.,
            Bowman,C., Watthey,L., Wallin,E., Hayea,W.S., Borodovsky,M.,
            Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 12901)
  AUTHORS   White,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12901
                     /organism="Helicobacter pylori 26695"
                     /strain="26695"
                     /db_xref="taxon:85962"
     gene            106. .1026
                     /gene="HP0160"
     CDS             106. .1026
                     /gene="HP0160"
                     /note="similar to GP:1786864 percent identity: 30.57;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical secreted protein"
                     /protein_id="AAD07230.1"
                     /db_xref="GI:2313248"
                     /translation="MIKSWTKKWFLILFLMASCSSYLVATTGEKYFKMATQAFKRGDY
                     HKAVAFYKRSCNLRVGVGCTSLGSMYEDGDGVDQNITKAVFYYRRGCNLRNHLACASL
                     GSMYEDGDGVQKNLPKAIYYYRRGCHLKGGVSCGSLGFMYFNGTGVKQNYAKALFLSK
                     YACSLNYGISCNFVGYMYRNAKGVQKDLKKALANFKRGCHLKDGASCVSLGYMYEVGM
                     DVKQNGEQALNLYKKGCYLKRGSGCHNVAVMYYTGKGVPKDLDKAISYYKKGCTLGFS
                     GSCKVLEEVIGKKSDDLQDDAQNDTQDDMQ"
     gene            1369. .1479
```

```
                     /gene="HP0161"
     CDS             1369. .1479
                     /gene="HP0161"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0161"
                     /protein_id="AAD07244.1"
                     /db_xref="GI:2313262"
                     /translation="MRTKLSPNISLGSPIALAFGRFTINIKELTTQAIIF"
     gene            complement(1928. .2650)
                     /gene="HP0162"
     CDS             complement(1928. .2650)
                     /gene="HP0162"
                     /note="similar to PID:1183839 percent identity: 36.65;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD07231.1"
                     /db_xref="GI:2313249"
                     /translation="MGRAFEYRRAAKEKRWDKMSKVFPKLAKAITLAAKDGGSEPDTN
                     AKLRTAILNAKAQNMPKDNIDAAIKRASSKEGNLSEITYEGKANPGVLIIMECMTDNP
                     TRTIANLKSYFNKTQGASIVPNGSLEFMFNRKSVFECLKNEVENLKLSLEDLEFALID
                     YGLEELEEVEDKIIIRGDYNSFKLLNEGFESLKLPILKASLQRIATTPIELNDEQMEL
                     TEKLLDRIEDDDDVVALYTNIE"
     gene            complement(2651. .3622)
                     /gene="HP0163"
     CDS             complement(2651. .3622)
                     /gene="HP0163"
                     /note="similar to GB:M57676 SP:P30950 PID:143039
                     GB:AL009126 percent identity: 50.47; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="delta-aminolevulinic acid dehydratase (hemB)"
                     /protein_id="AAD07232.1"
                     /db_xref="GI:2313250"
                     /translation="MFKRLRRLRSSENLRAMLRETRLNIDDFIAPLFVIESDSCIKNE
                     ISSMPGVYQMSMEPLLKECEELVGLGVKAVLLFGIPKHKDATGSHALNKDHIVAKAAK
                     EIKKRFKDLIVIADLCFCEYTDHGHCGILENASVSNDKTLEILNLQGLILAESGVDIL
                     APSNMMDGNVLSLRKTLDNAGYTHTPIMSYSTKFASSYYGPFRDVANSAPSFGDRKSY
                     QMDYANQKEALLESLEDEKQGADILMVKPALAYLDIVKEIRDHTLLPLALYNVSGEYA
                     MLKLAQKHNLINYESVLLETMTCFKRAGADMIISYHAKEVANLLQRN"
     gene            complement(3694. .4458)
                     /gene="HP0164"
     CDS             complement(3694. .4458)
                     /gene="HP0164"
                     /note="similar to PID:1064813 GB:AL009126 percent
                     identity: 28.20; identified by sequence similarity;
```



```
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="signal-transducing protein, histidine kinase"
                     /protein_id="AAD07233.1"
                     /db_xref="GI:2313251"
                     /translation="MSCKSKQKDEIGDLANEFDNCILKINAMNESRVLFLRSIMHELR
                     TPITKGKILSSMLKEELSCKRFSSIFDHLNMLIEQFARIEQLASKNYGSNKEKFLMSD
                     LIDKIEKMLLIDEDKESPIHVSSSNYIIEADFELFSIALKNMVDNAIKYSDDKQVFLD
                     FIGNNLVVSNKSKPLKEDFEKYLQPYFKSSNPSQAHGFGLGMYIIKNALEAMGLNLSY
                     HYSNGRICFTIHDCVFNSEYDLEEDNEELPPPPPKI"
     gene            complement(4455. .4976)
                     /gene="HP0165"
     CDS             complement(4455. .4976)
                     /gene="HP0165"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0165"
                     /protein_id="AAD07239.1"
                     /db_xref="GI:2313257"
                     /translation="MRFSIFFKVVALFMITLFSFGAFAYYFVSSQISHENYQNEMRHY
                     QFVTTINEILNNYSDYRAIEDYLYKIGFRETTIENLEKVLAKRRHQLHHRNIGYAEVF
                     KFSDMVFILLKKDEHFVLYKDLHSVSYRNYFLAITVGLLLILFLFLFVLQSLLPLREL
                     RSQVKPSLKGIKA"
     gene            complement(5002. .5679)
                     /gene="HP0166"
     CDS             complement(5002. .5679)
                     /gene="HP0166"
                     /note="similar to PID:881376 percent identity: 51.02;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="response regulator (ompR)"
                     /protein_id="AAD07234.1"
                     /db_xref="GI:2313252"
                     /translation="MIEVLMIEDDIELAEFLSEFLLQHGIHVTNYDEPYTGISAANTQ
                     NYDLLLLDLTLPNLDGLEVCRRISKQKHIPIIISGSARSDVEDKIKALDYGADDYLPKP
                     YDPKELLARIQSLLRRSHKKEEEVSEPGDANIFRVDKDSREVYMHEKKLDLTRAEYEIL
                     SLLISKKGYVFSRESIAIESESINPESSNKSIDVIIGRLRSKIEKNPKQPQYIISVRG
                     IGYKLEY"
     gene            complement(5915. .6388)
                     /gene="HP0167"
     CDS             complement(5915. .6388)
                     /gene="HP0167"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0167"
```



```
                     /protein_id="AAD07240.1"
                     /db_xref="GI:2313258"
                     /translation="MAGRMELKNIISETLNEIEKMAKTIDNNFDAAQKTPSFFKTPPY
                     LQNTLDPQNANVPLEPKNAAKIETQEKITEEKEEAPEIITEEIAQENPTQVLISNERV
                     FLKGLLERTLVLFKGMQALEEKEAMKRLDLVARFLQYQLSTLEKRLESLERENTE"
     gene            complement(6367. .6630)
                     /gene="HP0168"
     CDS             complement(6367. .6630)
                     /gene="HP0168"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. pylori predicted coding region HP0168"
                     /protein_id="AAD07241.1"
                     /db_xref="GI:2313259"
                     /translation="MPLETITLANIYEEQGFFEEALQIYINILKKTPNHAEALKQMKR
                     LEKIQKNGAPFKHDALLERHYLNFIKGDRLSVENLEKWLVEWN"
     gene            complement(6677. .7945)
                     /gene="HP0169"
     CDS             complement(6677. .7945)
                     /gene="HP0169"
                     /note="similar to GB:L42023 SP:P44700 PID:1003722
                     PID:1222349 PID:1204669 percent identity: 40.05;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="collagenase (prtC)"
                     /protein_id="AAD07235.1"
                     /db_xref="GI:2313253"
                     /translation="MNQVELLSPAGNLKKLKIALNYGADAVYGGVSHFSLRNRAGKEF
                     TLETFKEGIDYAHALNKKVYATINGFPPNSQLKLLEEHIDKMAELEPDAFIIAAPGVV
                     KLALKIAPHIPIHLSTQANVLNLLDAQVFYDLGVKRIVCARELSLNDAIEIKKALPNL
```

```
  Query Match             1.0%;  Score 61;  DB 1;  Length 12901;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-17;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    4159 agatatgatgtgtgagacctgtagggaatgcgttggagctcaaactctgtaaaatcccta 4218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1293 AGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCAAACTCTGTAAAATCCCTA 1352

Qy    4219 t 4219
           |
Db    1353 T 1353


RESULT 182
AF289393/c
LOCUS       AF289393      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori strain CR45 glutamate racemase (glr) and
```



```
            HP0519-like protein genes, partial cds.
ACCESSION   AF289393
VERSION     AF289393.1  GI:9964131
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /strain="CR45"
                     /db_xref="taxon:210"
                     /map="between HP0519 and HP0549"
                     /country="Costa Rica"
                     /note="cag pathogenicity island negative strain
                     isolated from patient with gastritis"
     gene            <1. .60
                     /gene="glr"
     CDS             <1. .60
                     /gene="glr"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG09836.1"
                     /db_xref="GI:9964132"
                     /translation="HASGDVIWLEKQAKEWLKL"
     CDS             72. .>291
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="HP0519-like protein"
                     /protein_id="AAG09837.1"
                     /db_xref="GI:9964133"
                     /translation="MFKDFYRTTLSFLKPLLLLVLLLPFSLCIADEYISISDDWDER
                     ARNQWDEIARNHKTYYFENGLDHFNQGQY"
BASE COUNT       91 a     45 c     60 g     95 t
ORIGIN
```

```
  Query Match             1.0%;  Score 60;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 4e-17;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy  5050 ttacaatttgagccattctttagcttgtttttctagccagatcacatcgccgctcgcatg 5109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 TTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCCGCTCGCATG 1


RESULT 183
AF289397/c
LOCUS       AF289397      291 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori strain CR10 glutamate racemase (glr) and
            HP0519-like protein genes, partial cds.
ACCESSION   AF289397
VERSION     AF289397.1  GI:9964143
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 291)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Helicobacter pylori"
                     /strain="CR10"
                     /db_xref="taxon:210"
                     /map="between HP0519 and HP0549"
                     /country="Costa Rica"
                     /note="cag pathogenicity island negative strain
                     isolated from patient with gastric adenocarcinoma"
     gene            <1. .60
                     /gene="glr"
     CDS             <1. .60
                     /gene="glr"
                     /codon_start=1
                     /transl_table=11
```



```
                     /product="glutamate racemase"
                     /protein_id="AAG09844.1"
                     /db_xref="GI:9964144"
                     /translation="HASGDVIWLEKQAKEWLKL"
     CDS             72. .>291
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="HP0519-like protein"
                     /protein_id="AAG09845.1"
                     /db_xref="GI:9964145"
                     /translation="MFKDFYRTTLSFLKPLLLLLVLLLPFSLCIADEYISISDDWDER
                     ARNQWDEIARNHKTYYFENGLDHFNQGQY"
BASE COUNT      91 a     44 c     61 g     95 t
ORIGIN


  Query Match             1.0%;  Score 60;  DB 2;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 4e-17;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy  5050 ttacaatttgagccattctttagcttgtttttctagccagatcacatcgccgctcgcatg 5109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 TTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCCGCTCGCATG 1


RESULT 184
HPY269853
LOCUS       HPY269853     464 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            HK43 from China: Hong Kong.
ACCESSION   AJ269853
VERSION     AJ269853.1  GI:5921283
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 464)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
```



```
     source          1. .464
                     /organism="Helicobacter pylori"
                     /isolate="HK43"
                     /db_xref="taxon:210"
                     /country="China: Hong Kong"
     gene            1. .464
                     /gene="cagA"
     CDS             1. .>464
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56426.1"
                     /db_xref="GI:5921284"
                     /translation="MTNETIVQTTTLDQTPNQTDFVPQRFINNLQVAPIKVDNAVASF
                     DPDQKPIVDKNDRDNRQAFEKISQLREEYANKAIKNPTKKNQYFSDFINKSNDLINKD
                     NLIAVDSSVDSFKKFGDQRYQIFTSWVSLQKDPSQINTRQIRNFMENIIQPP"
BASE COUNT     175 a     95 c     70 g    124 t
ORIGIN


  Query Match             1.0%;  Score 59;  DB 2;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-16;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   652 gatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttga 710
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 GATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGA 191


RESULT 185
HPY269874
LOCUS       HPY269874     464 bp    DNA             BCT       23-SEP-1999
DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
            US258 from United States.
ACCESSION   AJ269874
VERSION     AJ269874.1  GI:5921325
KEYWORDS    cagA gene; CagA protein.
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
  TITLE     Distinct variants of Helicobacter pylori cagA are associated with
            vacA subtypes
  JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
  MEDLINE   99294773
REFERENCE   2  (bases 1 to 464)
  AUTHORS   van Doorn,L.J.
  TITLE     Direct Submission
```



```
  JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
            Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Helicobacter pylori"
                     /isolate="US258"
                     /db_xref="taxon:210"
                     /country="USA"
     gene            1. .464
                     /gene="cagA"
     CDS             1. .>464
                     /gene="cagA"
                     /codon_start=1
                     /transl_table=11
                     /product="CagA protein"
                     /protein_id="CAB56485.1"
                     /db_xref="GI:5921326"
                     /translation="MTNETIDQTRTPDQTQSQTAFDPQRFINNLQVAFLKVDSAVASL
                     DPDQKPIVDKNDRDNRQAFEKISQLREEFANKAIKNPTKKNQYFSNFISKSNDLINKD
                     GLIDTGSSIKSFQKFGDQRYRIFTSWVVHQNDPSKINTRSIRNFMEHTIQPP"
BASE COUNT     171 a     96 c     77 g    120 t
ORIGIN


  Query Match             1.0%;  Score 59;  DB 2;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-16;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   652 gatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttga 710
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 GATCCTGATCAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGA 191


RESULT 186
AF191007
LOCUS       AF191007      387 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain PJAP59A cag pathogenicity island
            polymorphic right end, type II motif right-end.
ACCESSION   AF191007
VERSION     AF191007.1  GI:10567570
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
```

```
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 387)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .387
                     /organism="Helicobacter pylori"
                     /strain="PJAP59A"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="Peru"
                     /note="isolated from ethnic Japanese; polymorphic region
                     at right end of the cag pathogenicity island; type II
                     motif right-end; similar to Helicobacter pylori strain Hpl
                     type II motif presented in GenBank Accession Number
                     AF190661"
BASE COUNT      104 a     70 c     74 g    139 t
ORIGIN


  Query Match             1.0%;  Score 57;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-15;
  Matches   57;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    4949 atctctctaaagcgctttataatcaacacaatacccttatagtgtgagctatagccc 5005
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82 ATCTCTCTAAAGCGCTTTATAATCAACACAATACCCTTATAGTGTGAGCTATAGCCC 138


RESULT 187
AF191003
LOCUS       AF191003      444 bp    DNA             BCT       04-OCT-2000
DEFINITION  Helicobacter pylori strain HU-TK cag pathogenicity island
            polymorphic right end, type II motif.
ACCESSION   AF191003
VERSION     AF191003.1  GI:10567566
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
```



```
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torres,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 444)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .444
                     /organism="Helicobacter pylori"
                     /strain="HU-TK"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /note="isolated from Japanese gastric ulcer patient;
                     polymorphic region at right end of the cag pathogenicity
                     island; type II motif; similar to Helicobacter pylori
                     strain Hpl type II motif presented in GenBank Accession
                     Number AF190661"
BASE COUNT      128 a     87 c     76 g    153 t
ORIGIN


  Query Match             1.0%;  Score 57;  DB 1;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-15;
  Matches   57;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    4949 atctctctaaagcgctttataatcaacacaatacccttatagtgtgagctatagccc 5005
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     353 ATCTCTCTAAAGCGCTTTATAATCAACACAATACCCTTATAGTGTGAGCTATAGCCC 409


RESULT 188
AF289454
LOCUS       AF289454      513 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR50 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289454
VERSION     AF289454.1  GI:9964262
KEYWORDS    .
```



```
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 513)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Helicobacter pylori"
                     /isolate="CR50"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>513
                     /gene="cagA"
     CDS             <1. .>513
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09903.1"
                     /db_xref="GI:9964263"
                     /translation="MFSDIKKELNAKLFGNFNNNNNNGLKNSTEPIYAKVNKKKTGQV
                     ASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDK
                     VDDLSKVGLSREQQLKQKIDNLSQAVSEAKAGFFGNLEQTIDKLKDSTKNNNVNLWAE
                     SVKKVPASLSA"
     misc_feature    76. .99
                     /gene="cagA"
                     /note="Region: tyrosine phosphorylation motif B"
BASE COUNT      187 a     89 c    106 g    131 t
ORIGIN


  Query Match             1.0%;  Score 57;  DB 2;  Length 513;
  Best Local Similarity   100.0%;  Pred. No. 1e-15;
  Matches   57;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3358 ttccctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaagg 3414
```



```
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     283 TTCCCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGG 339


RESULT 189
AF289452
LOCUS       AF289452      567 bp    DNA             BCT       02-SEP-2000
DEFINITION  Helicobacter pylori isolate CR48 cytotoxin associated protein A
            (cagA) gene, partial cds.
ACCESSION   AF289452
VERSION     AF289452.1  GI:9964258
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Composition and gene expression of the cag pathogenicity island in
            Helicobacter pylori strains isolated from gastric carcinoma and
            gastritis patients in Costa Rica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 567)
  AUTHORS   Occhialini,A., Marais,A., Urdaci,M., Sierra,R., Munoz,N.,
            Covacci,A. and Megraud,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Laboratory of Bacteriology, University of
            Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Helicobacter pylori"
                     /isolate="CR48"
                     /db_xref="taxon:210"
                     /country="Costa Rica"
                     /note="isolated from a patient with chronic
                     gastritis-associated dyspepsia"
     gene            <1. .>567
                     /gene="cagA"
     CDS             <1. .>567
                     /gene="cagA"
                     /note="CagA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytotoxin associated protein A"
                     /protein_id="AAG09901.1"
                     /db_xref="GI:9964259"
                     /translation="MFSDIKKELNAKLFGNFNNNNNNGLKNEPIYAKVNKKKTGQVAS
                     PEEPIYAQVAKKVTGKIDQLNQAASGPCGVGQAGFPLKRHDKVDDLSKVGRSVSPEPI
                     YATIDDLGGPFPLKRHDKVDDLSKVGLSREQQLKQKIDNLSQAVSEAKAGFFGNLEQT
                     INKLKDSTKNNTVNLWAESAKKVPASLSA"
```

```
BASE COUNT      208 a     99 c    121 g    139 t
ORIGIN


  Query Match             1.0%;  Score 57;  DB 2;  Length 567;
  Best Local Similarity   100.0%;  Pred. No. 1e-15;
  Matches   57;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3358 ttccctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaagg 3414
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 TTCCCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGG 393


RESULT 190
AF198480
LOCUS       AF198480      220 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain PeruGJ6 cytotoxin associated protein
            CagA (cagA) gene, partial cds.
ACCESSION   AF198480
VERSION     AF198480.1  GI:7248686
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
            Xiao,S.D., Tytgat,G.N. and Dankert,J.
  TITLE     cagA-positive Helicobacter pylori populations in China and The
            Netherlands are distinct
  JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
  MEDLINE   98234000
   PUBMED   9573056
REFERENCE   2  (bases 1 to 220)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C.Y., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L.,
            Torres,O., Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   3  (bases 1 to 220)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
```



```
            USA
FEATURES            Location/Qualifiers
     source         1. .220
                    /organism="Helicobacter pylori"
                    /strain="PeruGJ6"
                    /db_xref="taxon:210"
                    /note="European type; isolated from gastric juice sample
                    from native Peruvian"
     gene           <1. .>220
                    /gene="cagA"
     CDS            <1. .>220
                    /gene="cagA"
                    /note="similar to Helicobacter pylori strain 26695 HP0547"
                    /codon_start=2
                    /transl_table=11
                    /product="cytotoxin associated protein CagA"
                    /protein_id="AAF43620.1"
                    /db_xref="GI:7248687"
                    /translation="RDQQGNNVATIINVHMKNGSGLVIAGGEKGINNPSFYLYKEDQL
                    TGSQRALSQEEIRNKIDFMEFLAQNNAKL"
BASE COUNT       87 a     37 c     44 g     52 t
ORIGIN


  Query Match             0.9%;  Score 56;  DB 1;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-15;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1642 aaccctagtttttatctctacaaagaagaccaactcacaggctcacaacgagcatt 1697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     98 AACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATT 153


RESULT 191
AF202225
LOCUS       AF202225      286 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain India19A CagA (cagA) gene, partial cds.
ACCESSION   AF202225
VERSION     AF202225.1  GI:7229441
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
            Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
            Nair,G.B. and Berg,D.E.
  TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
            India
  JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
```



```
  MEDLINE   20270153
   PUBMED   10809703
REFERENCE   2  (bases 1 to 286)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Chowdhury,A., Datta,S., Nair,G.B.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) Molecular Microbiology, Washington
            University, 660 South Euclid Ave., St. Louis, MO 63110, USA
FEATURES            Location/Qualifiers
     source         1. .286
                    /organism="Helicobacter pylori"
                    /strain="India19A"
                    /db_xref="taxon:210"
     gene           <1. .>286
                    /gene="cagA"
     CDS            <1. .>286
                    /gene="cagA"
                    /note="similar to Helicobacter pylori 26695 HP0547"
                    /codon_start=1
                    /transl_table=11
                    /product="CagA"
                    /protein_id="AAF42827.1"
                    /db_xref="GI:7229442"
                    /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                    EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
BASE COUNT      103 a     52 c     67 g     64 t
ORIGIN


  Query Match             0.9%;  Score 56;  DB 1;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-15;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1642 aaccctagtttttatctctacaaagaagaccaactcacaggctcacaacgagcatt 1697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    148 AACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATT 203


RESULT 192
AF202219
LOCUS       AF202219      287 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain India3B CagA (cagA) gene, partial cds.
ACCESSION   AF202219
VERSION     AF202219.1  GI:7229429
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
```



```
            Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
            Nair,G.B. and Berg,D.E.
  TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
            India
  JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
  MEDLINE   20270153
   PUBMED   10809703
REFERENCE   2  (bases 1 to 287)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Chowdhury,A., Datta,S., Nair,G.B.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) Molecular Microbiology, Washington
            University, 660 South Euclid Ave., St. Louis, MO 63110, USA
FEATURES            Location/Qualifiers
     source         1. .287
                    /organism="Helicobacter pylori"
                    /strain="India3B"
                    /db_xref="taxon:210"
                    /note="from gastric ulcer patient from Calcutta, India"
     gene           <1. .>287
                    /gene="cagA"
     CDS            <1. .>287
                    /gene="cagA"
                    /note="similar to Helicobacter pylori 26695 HP0547"
                    /codon_start=2
                    /transl_table=11
                    /product="CagA"
                    /protein_id="AAF42821.1"
                    /db_xref="GI:7229430"
                    /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                    EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLGNLSE"
BASE COUNT      102 a     51 c     68 g     66 t
ORIGIN


  Query Match             0.9%;  Score 56;  DB 1;  Length 287;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-15;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1642 aaccctagtttttatctctacaaagaagaccaactcacaggctcacaacgagcatt 1697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 AACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATT 204


RESULT 193
AF202222
LOCUS       AF202222      287 bp    DNA             BCT       19-MAY-2000
DEFINITION  Helicobacter pylori strain India10A CagA (cagA) gene, partial cds.
ACCESSION   AF202222
VERSION     AF202222.1  GI:7229435
KEYWORDS    .
```

```
  SOURCE      Helicobacter pylori.
    ORGANISM  Helicobacter pylori
              Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
              Helicobacter.
  REFERENCE   1  (bases 1 to 287)
    AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Jeong,J.Y., Datta,S., Ito,Y.,
              Chowdhury,A., Chowdhury,S., Santra,A., Bhattacharya,S.K., Azuma,T.,
              Nair,G.B. and Berg,D.E.
    TITLE     Distinctiveness of genotypes of Helicobacter pylori in Calcutta,
              India
    JOURNAL   J. Bacteriol. 182 (11), 3219-3227 (2000)
    MEDLINE   20270153
     PUBMED   10809703
  REFERENCE   2  (bases 1 to 287)
    AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Chowdhury,A., Datta,S., Nair,G.B.
              and Berg,D.E.
    TITLE     Direct Submission
    JOURNAL   Submitted (08-NOV-1999) Molecular Microbiology, Washington
              University, 660 South Euclid Ave., St. Louis, MO 63110, USA
  FEATURES             Location/Qualifiers
       source          1. .287
                       /organism="Helicobacter pylori"
                       /strain="India10A"
                       /db_xref="taxon:210"
       gene            <1. .>287
                       /gene="cagA"
       CDS             <1. .>287
                       /gene="cagA"
                       /note="similar to Helicobacter pylori 26695 HP0547"
                       /codon_start=2
                       /transl_table=11
                       /product="CagA"
                       /protein_id="AAF42824.1"
                       /db_xref="GI:7229436"
                       /translation="GNGGPGARHDWNATVGYKDQQGNNVATIINVHMKNGSGLVIAGG
                       EKGINNPSFYLYKEDQLTGSQRALSQEEIRNKIDFMEFLAQNNAKLDNLSE"
  BASE COUNT      103 a      52 c      67 g      65 t
  ORIGIN
```

```
  Query Match           0.9%;  Score 56;  DB 1;  Length 287;
  Best Local Similarity  100.0%;  Pred. No. 3.2e-15;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1642 aaccctagttttatctctacaaagaagaccaactcacaggctcacaacgagcatt 1697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 AACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACAGGCTCACAACGAGCATT 204
```

```
  RESULT 194
  HPY269868
```



```
  LOCUS       HPY269868     452 bp    DNA             BCT       23-SEP-1999
  DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate
              JA147 from Japan.
  ACCESSION   AJ269868
  VERSION     AJ269868.1  GI:5921313
  KEYWORDS    cagA gene; CagA protein.
  SOURCE      Helicobacter pylori.
    ORGANISM  Helicobacter pylori
              Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
              Helicobacter.
  REFERENCE   1  (bases 1 to 452)
    AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
    TITLE     Distinct variants of Helicobacter pylori cagA are associated with
              vacA subtypes
    JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
    MEDLINE   99294773
  REFERENCE   2  (bases 1 to 452)
    AUTHORS   van Doorn,L.J.
    TITLE     Direct Submission
    JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
              Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
              NETHERLANDS
  FEATURES             Location/Qualifiers
       source          1. .452
                       /organism="Helicobacter pylori"
                       /isolate="JA147"
                       /db_xref="taxon:210"
                       /country="Japan"
       gene            1. .452
                       /gene="cagA"
       CDS             1. .>452
                       /gene="cagA"
                       /codon_start=1
                       /transl_table=11
                       /product="CagA protein"
                       /protein_id="CAB56441.1"
                       /db_xref="GI:5921314"
                       /translation="MTNETIDQTTTPDRTDFVPQRFINNLQVAFIKVDNAVASFDPDQ
                       KPIVDKNDRDNRQAFEKISQLREEYANKAIKNPTKKNQYFSDFINKSNDLINKDNLIA
                       VDSSVESFRKFGDQRYQIFTSWVSLQKDPSKINTQQIRNFMENIIQPP"
  BASE COUNT      170 a      90 c      70 g     122 t
  ORIGIN
```

```
  Query Match           0.9%;  Score 56;  DB 2;  Length 452;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-15;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    763 accaaaaagaatcagtatttttcagactttatcaataagagcaatgatttaatcaa 818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    232 ACCAAAAAGAATCAGTATTTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAA 287
```



```
  RESULT 195
  HPY269856
  LOCUS       HPY269856     464 bp    DNA             BCT       23-SEP-1999
  DEFINITION  Helicobacter pylori partial cagA gene for CagA protein, isolate HK8
              from China: Hong Kong.
  ACCESSION   AJ269856
  VERSION     AJ269856.1  GI:5921289
  KEYWORDS    cagA gene; CagA protein.
  SOURCE      Helicobacter pylori.
    ORGANISM  Helicobacter pylori
              Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
              Helicobacter.
  REFERENCE   1  (bases 1 to 464)
    AUTHORS   van Doorn,L.J., Figueiredo,C., Sanna,R., Blaser,M.J. and Quint,W.G.
    TITLE     Distinct variants of Helicobacter pylori cagA are associated with
              vacA subtypes
    JOURNAL   J. Clin. Microbiol. 37 (7), 2306-2311 (1999)
    MEDLINE   99294773
  REFERENCE   2  (bases 1 to 464)
    AUTHORS   van Doorn,L.J.
    TITLE     Direct Submission
    JOURNAL   Submitted (13-SEP-1999) van Doorn L.J., Molecular Biology, Delft
              Diagnostic Laboratory, R. de Graafweg 7, Delft, 2625 AD,
              NETHERLANDS
  FEATURES             Location/Qualifiers
       source          1. .464
                       /organism="Helicobacter pylori"
                       /isolate="HK8"
                       /db_xref="taxon:210"
                       /country="China: Hong Kong"
       gene            1. .464
                       /gene="cagA"
       CDS             <1. .>464
                       /gene="cagA"
                       /codon_start=1
                       /transl_table=11
                       /product="CagA protein"
                       /protein_id="CAB56429.1"
                       /db_xref="GI:5921290"
                       /translation="MTNETIDQTTTPDQTPSQTDFVPQRFINNLQVAFLKVDNAVASF
                       DPDQKPIVDKNDRDNRQAFEKISQLREEYANKAIKNPAKKNQYFSDFINKTNDLINKD
                       NLIAVDSSVDSFKKFGDQRYQIFTSWVSLQKDPSKINTQQIRNFMENIIQPP"
  BASE COUNT      175 a      96 c      70 g     123 t
  ORIGIN
```

```
  Query Match           0.9%;  Score 56;  DB 2;  Length 464;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-15;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy    594 atttatcaataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcat 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 ATTTATCAATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCAT 130
```

```
  RESULT 196
  HPY252967
  LOCUS       HPY252967     244 bp    DNA             BCT       08-SEP-1999
  DEFINITION  Helicobacter pylori partial cagA gene for cytotoxin associated
              protein A.
  ACCESSION   AJ252967
  VERSION     AJ252967.1  GI:5851965
  KEYWORDS    cagA gene; cytotoxin associated protein A.
  SOURCE      Helicobacter pylori.
    ORGANISM  Helicobacter pylori
              Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
              Helicobacter.
  REFERENCE   1  (bases 1 to 244)
    AUTHORS   van der Ende,A., Pan,Z.J., Bart,A., van der Hulst,R.W., Feller,M.,
              Xiao,S.D., Tytgat,G.N. and Dankert,J.
    TITLE     cagA-positive Helicobacter pylori populations in China and The
              Netherlands are distinct
    JOURNAL   Infect. Immun. 66 (5), 1822-1826 (1998)
    MEDLINE   98234000
  REFERENCE   2  (bases 1 to 244)
    AUTHORS   van der Ende,A.
    TITLE     Direct Submission
    JOURNAL   Submitted (07-SEP-1999) van der Ende A., Medical Microbiology,
              Academic Medical Center, Meibergdreef 15, Amsterdam, 1105 AZ,
              NETHERLANDS
  FEATURES             Location/Qualifiers
       source          1. .244
                       /organism="Helicobacter pylori"
                       /strain="Dutch 239"
                       /db_xref="taxon:210"
                       /country="Netherlands"
       gene            1. .244
                       /gene="cagA"
       CDS             <1. .>244
                       /gene="cagA"
                       /codon_start=1
                       /transl_table=11
                       /product="cytotoxin associated protein A"
                       /protein_id="CAB55443.1"
                       /db_xref="GI:5851965"
                       /translation="VGYKDQQGNNVATIINAHMKNGSGLVIAGGEKGINDPSPYLYKE
                       DQLTGSQRALSQEEIQNKVDFMEFLAQNNAKLDNLSE"
  BASE COUNT       94 a      42 c      52 g      56 t
  ORIGIN
```

```
  Query Match             0.9%;  Score 55;  DB 2;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-15;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1587 gcatatgaaaaacggcagtggcttagtcatagcaggtggtgagaaagggattaac 1641
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51 GCATATGAAAAACGGCAGTGGCTTAGTCATAGCAGGTGGTGAGAAAGGGATTAAC 105


RESULT 197
AF201075
LOCUS       AF201075      427 bp    DNA             BCT       14-SEP-2000
DEFINITION  Helicobacter pylori strain India66A cag pathogenicity island
            polymorphic right end, type V motif and glutamate racemase (glr)
            gene, partial cds.
ACCESSION   AF201075
VERSION     AF201075.1  GI:10121333
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Kersulyte,D., Mukhopadhyay,A.K., Velapatino,B., Su,W.W., Pan,Z.J.,
            Garcia,C., Hernandez,V., Valdez,Y., Mistry,R.S., Gilman,R.H.,
            Yuan,Y., Gao,H., Alarcon,T., Lopez-Brea,M., Balakrish Nair,G.,
            Chowdhury,A., Datta,S., Shirai,M., Nakazawa,T., Ally,R., Segal,I.,
            Wong,B.C., Lam,S.K., Olfat,F.O., Boren,T., Engstrand,L., Torrea,O.,
            Schneider,R., Thomas,J.E., Czinn,S. and Berg,D.E.
  TITLE     Differences in genotypes of Helicobacter pylori from different
            human populations
  JOURNAL   J. Bacteriol. 182 (11), 3210-3218 (2000)
  MEDLINE   20270152
   PUBMED   10809702
REFERENCE   2  (bases 1 to 427)
  AUTHORS   Mukhopadhyay,A.K., Kersulyte,D., Velapatino,B., Su,W.W., Pan,Z.J.
            and Berg,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1999) Molecular Microbiology, Washington
            University Medical School, 4566 Scott Ave., St. Louis, MO 63110,
            USA
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Helicobacter pylori"
                     /strain="India66A"
                     /db_xref="taxon:210"
                     /map="between HP0547 and HP0549 in 26695 genome"
                     /country="India: Calcutta"
                     /note="polymorphic region at right end of the cag
                     pathogenicity island; isolated from gastric ulcer patient"
```



```
     misc_feature    1. .197
                     /note="type V motif"
     misc_feature    29. .149
                     /note="similar to Helicobacter pylori insertion sequence
                     IS606 presented in GenBank Accession Number U95957"
     variation       150. .191
                     /note="absent in type I, II, III and IV motifs"
                     /replace=""
     repeat_region   197. .427
                     /note="defines the terminus of the cag pathogenicity
                     island"
                     /rpt_type=direct
     gene            complement(197. .>427)
                     /gene="glr"
                     /note="flanks cag pathogenicity island"
     CDS             complement(197. .>427)
                     /gene="glr"
                     /note="similar to Helicobacter pylori strain 26695
                     glutamate racemase HP0549"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate racemase"
                     /protein_id="AAG13133.1"
                     /db_xref="GI:10121334"
                     /translation="GCTHPPLIAQKIEGYFMEHPALPTPPLLIHSGDAIVGYLQQKYA
                     LKKNACAPPKVEFHASGDVIWLEKQAKEWLKL"
BASE COUNT      131 a     89 c     80 g    127 t
ORIGIN


  Query Match             0.9%;  Score 55;  DB 1;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-15;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4328 actaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaaaa 4382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 ACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAAAA 136


RESULT 198
AF049648
LOCUS       AF049648     4735 bp    DNA             BCT       13-JAN-1999
DEFINITION  Helicobacter pylori strain F68 vacuolating cytotoxin precursor
            (vacA) gene, complete cds; and FepC (fepC) gene, partial cds.
ACCESSION   AF049648
VERSION     AF049648.1  GI:4151177
KEYWORDS    .
SOURCE      Helicobacter pylori.
  ORGANISM  Helicobacter pylori
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
```



```
REFERENCE   1  (bases 1 to 4735)
  AUTHORS   Ito,Y., Azuma,T., Ito,S., Suto,H., Miyaji,H., Yamazaki,Y., Kohli,Y.
            and Kuriyama,M.
  TITLE     Full-length sequence analysis of the vacA gene from cytotoxic and
            noncytotoxic Helicobacter pylori
  JOURNAL   J. Infect. Dis. 178 (5), 1391-1398 (1998)
  MEDLINE   98453456
REFERENCE   2  (bases 1 to 4735)
  AUTHORS   Ito,Y., Azuma,T., Ito,S., Suto,H., Miyaji,H., Yamazaki,Y., Kohli,Y.
            and Kuriyama,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1998) Second Department of Internal Medicine,
            Fukui Medical University, Matsuoka-cho, Yoshida-gun, Fukui
            910-1193, Japan
FEATURES             Location/Qualifiers
     source          1. .4735
                     /organism="Helicobacter pylori"
                     /strain="F68"
                     /db_xref="taxon:210"
     gene            275. .4165
                     /gene="vacA"
     CDS             275. .4165
                     /gene="vacA"
                     /codon_start=1
                     /transl_table=11
                     /product="vacuolating cytotoxin precursor"
                     /protein_id="AAD04283.1"
                     /db_xref="GI:4151178"
                     /translation="MELQQTHRKINRPLVSLALVGLLVSITPQKSHAAPFTTVIIPAI
                     VGGIATGAAVGTVSGLLGWGLKQAEEANKTPDKPDKVWRIQAGRGFNEPPNKEYDLYK
                     SLLSSKIDGGWDWGNAARHYWVKGGQWNKLEVDMKDAVGTYKLSGLRNFTGGDLDVNH
                     QKATLRLGQFNGNSFTSFKDGTNRTTRVDFNAKNILIDNFLEINNRVGSGAGRKASST
                     VLTLQASEGITSGKNAEISLYDGATLNLASNSVKLMGNVWMGRLQYVGAYLAPSYSTI
                     NTSKVTGEVNFNHLTVGDRNAAQAGIIASKKTYIGTDLGWGSAGLNIIAPPEGGYKDK
                     PNNTNSQSGAKSDKNESAKNDKQESSQNNSNTQVINPPNSGQKTEIQPTQVIDGPPAG
                     AKDTMVNINRINTNADGTIKVGGYTASLTTNAAHLNIGKGGVNLSNQASGRTLLVENL
                     TGNITVEGALRVNNQVGGYALAGSSANFEPKAGVDTKNGTIAPNNNISLGRFVNLKAS
                     AHTVNFKDIDTGNGGFNTLDFSGVTNKVNINKLITASTNVAIKNFNINELLVKTNGIS
                     VGEYTNFSEDIGNQSRINTVRLETGTRSIYSGGVKFKGGEKLVINDFYYAPWNYFDAR
                     NIKNVEITNKLAFGPQGSPWGTAKLMFNNLTLGPNAVMDYSQFSNVTIQGNFVNNQGT
                     INYLVRGGNIETLSVGNAAVMSFNNDIDSATGFYKPLIKINSAQDLIKNKEHVLLKAK
                     IIGYGNVSLGTNSISNANLIEQFNERLALYNNNRMDTCVVRNTDDIKACGMAIGDQA
                     MVNNPDNYKYLIGKAWKNIGISKTANGSKISVRYLGNATPPENGGNTTNLPTNTTNNA
                     RFASYALIKNAPFAQTNATPNLVAINKHNPGTIESVPELANRSKDIDTLYAHSGAQGR
                     DLLQTLLIDSHNAGYARTMIDATSANEITKQLNAANSALNNIASLEHKTNGLQTLSLS
                     NAMILNSRLVNLSRRHTNNIDSFAQRLQALKDQRFASLESAAEVLYQFAPKYEKPTNV
                     WANAIGGASLNSGGNTSLYGTSAGVDAYLNEKVEAIVGGPGSYGYSSFSNQANSLNSG
                     ANNANFGVYSRIFANNRHEFDFEAQGAVGSDQSSLNFKSALLRDFNQSYNYLAYGAATR
                     ASYGYDFAFFRNALVLKPSVGVSYNHLGSTNFESNSTHKTALKNGASSQHLFNASANV
                     EARYYYGDTSYFYMNAGVLQEFANFGSSNAVSLNTFKVNTARNPLNTHARVMMGGELK
                     LAKEVFLNLGFIYLHNLISNIGHFASNLGMRYSF"
```



```
     repeat_region   4222. .4490
                     /note="mini IS605"
     gene            complement(4494. .>4735)
                     /gene="fepC"
     CDS             complement(4494. .>4735)
                     /gene="fepC"
                     /codon_start=3
                     /transl_table=11
                     /product="FepC"
                     /protein_id="AAD04284.1"
                     /db_xref="GI:4151179"
                     /translation="LFFDAIKDEMKKRELSVLVNIHDPNLVARHSTHVVMLKDKKLFL
                     QASTQIAMTSHNLSVLYDTPLEAIWHDNKLVVYAL"
BASE COUNT     1488 a     933 c    1009 g    1305 t
ORIGIN


  Query Match             0.9%;  Score 55;  DB 1;  Length 4735;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-15;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4142 caaagttacgccttggagatatgatgtgtgagacctgtagggaatgcgttggag 4196
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4324 CAAAGTTACGCCTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAG 4378


RESULT 199
AE000570
LOCUS       AE000570    10482 bp    DNA             BCT       06-APR-1999
DEFINITION  Helicobacter pylori 26695 section 48 of 134 of the complete genome.
ACCESSION   AE000570 AE000511
VERSION     AE000570.1  GI:2313672
KEYWORDS    .
SOURCE      Helicobacter pylori 26695.
  ORGANISM  Helicobacter pylori 26695
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 10482)
  AUTHORS   Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and
            Venter,J.C.
  TITLE     The complete genome sequence of the gastric pathogen Helicobacter
            pylori
  JOURNAL   Nature 388 (6642), 539-547 (1997)
  MEDLINE   97394467
  REMARK    Erratum:[[published erratum appears in Nature 1997 Sep
```

```
            25;389(6649):412)]
  REFERENCE 2  (bases 1 to 10482)
    AUTHORS Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
            Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
            Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
            Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
            Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D.,
            Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
            Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C.,
            Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M.,
            Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
    TITLE   Direct Submission
    JOURNAL Submitted (06-AUG-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REFERENCE 3  (bases 1 to 10482)
    AUTHORS White,O.
    TITLE   Direct Submission
    JOURNAL Submitted (17-MAR-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  FEATURES             Location/Qualifiers
     source            1. .10482
                       /organism="Helicobacter pylori 26695"
                       /strain="26695"
                       /db_xref="taxon:85962"
     gene              complement(67. .504)
                       /gene="HP0556"
     CDS               complement(67. .504)
                       /gene="HP0556"
                       /note="hypothetical protein; identified by GeneMark;
                       putative"
                       /codon_start=1
                       /transl_table=11
                       /product="H. pylori predicted coding region HP0556"
                       /protein_id="AAD07631.1"
                       /db_xref="GI:2313682"
                       /translation="MVTHEKIKSRFSRNWSLRNRGRHFASASVYFPSLLVITAVNRSS
                       AVAWLLMPEHLIGWFLISPSGEFVADMAPGKKSKIFKTRPGISIVSGVSLLLGALPAH
                       LFFVWFGFINWWAVLFIEAGAGLLVGCVIQKIFFGKYWVDRYY"
     gene              complement(973. .1911)
                       /gene="HP0557"
     CDS               complement(973. .1911)
                       /gene="HP0557"
                       /note="similar to GB:M96394 SP:P30867 PID:1208956
                       PID:147322 PID:1122205 percent identity: 50.33; identified
                       by sequence similarity; putative"
                       /codon_start=1
                       /transl_table=11
                       /product="acetyl-coenzyme A carboxylase (accA)"
                       /protein_id="AAD07624.1"
                       /db_xref="GI:2313675"
                       /translation="MAVYLDFENHIKEIQNEIELALIRGDEDAKEILEKRLDKEVKSI
```



```
                       YSNLTDFQKLQLARHPDRPYAMDYIDLILKDKYEVFGDRHYNDDKAIVCFVGKIDNVP
                       VVVIGEEKGRGTKNKLLRNFGMPNPCGYRKALKMAKPAEKFNLPILMLVDTAGAYPGI
                       GAEERGQSEAIAKNLQEPASLKVPTISVIIGEGGSGGALAIAVADKLAMMEYSIFSVI
                       SPEGCAAILWKDPSKTEVAIKAMKITPRDLKEAGLIDDIILEPSKGAHRDKFSAANTI
                       KEYFLDALRTIQQDPHFLDNRYQKLMSLGSFVESMN"
     gene              complement(1934. .3172)
                       /gene="HP0558"
     CDS               complement(1934. .3172)
                       /gene="HP0558"
                       /note="similar to GB:Z34979 SP:P39435 GB:U20767 PID:572680
                       PID:664870 percent identity: 50.00; identified by sequence
                       similarity; putative"
                       /codon_start=1
                       /transl_table=11
                       /product="beta ketoacyl-acyl carrier protein synthase II
                       (fabF)"
                       /protein_id="AAD07625.1"
                       /db_xref="GI:2313676"
                       /translation="MRRIVVTGMGMINSLGLNKEDSFLAIAKGECGIKHIESFDASAF
                       PVRIAGEITDFDPTEVMNPKDVKKAGRFIQLALKATREAMKDSGILDAHNRCPEELAN
                       RMGVSSGSGIGGLGNIEANSIFCFEKGPRKVNPFFITSALVNMIGGFTSIEFGIKGPN
                       LSSVTACAAGTHAIIEAVKTILLNGADKMLVVGAESTICPVGIGGFASIKALSTRNDD
                       PKKASRPFDKDRNGFVMGEGAGALVLEEYESAKRRGAKIYAEFAGYGESGDANHITAP
                       APEGEGAFRAMKMALEMAKVEIGYVNAHGTSTHYNDLYESIALKNVFGSKEKVPPVSS
                       TKGQIGHCLGAAGALEAVISIMAMNQGILPPTINQEMPDPECDLDIPNAAREKRVDA
                       VMSNSFGFGGTNGVVIFKKA"
     gene              complement(3480. .3716)
                       /gene="HP0559"
     CDS               complement(3480. .3716)
                       /gene="HP0559"
                       /note="similar to GB:L42023 SP:P43709 PID:1003209
                       PID:1222068 PID:1204411 percent identity: 55.26;
                       identified by sequence similarity; putative"
                       /codon_start=1
                       /transl_table=11
                       /product="acyl carrier protein (acpP)"
                       /protein_id="AAD07626.1"
                       /db_xref="GI:2313677"
                       /translation="MALFEDIQAVIAEQLNVDAVQVTPEAEFVKDLGADSLDVVELIM
                       ALEEKFGVEIPDEQAEKIINVGDVVKYIEDNKLA"
     gene              complement(3769. .3849)
                       /gene="HP0560"
     CDS               complement(3769. .3849)
                       /gene="HP0560"
                       /note="hypothetical protein; identified by GeneMark;
                       putative"
                       /codon_start=1
                       /transl_table=11
                       /product="H. pylori predicted coding region HP0560"
                       /protein_id="AAD07633.1"
                       /db_xref="GI:2313684"
```



```
                       /translation="MGIIYLILFLIVIYLLYRILDVLEQK"
     gene              complement(3947. .4690)
                       /gene="HP0561"
     CDS               complement(3947. .4690)
                       /gene="HP0561"
                       /note="similar to GB:M84991 SP:P25716 PID:145881
                       PID:145888 GB:U00096 percent identity: 45.71; identified
                       by sequence similarity; putative"
                       /codon_start=1
                       /transl_table=11
                       /product="3-ketoacyl-acyl carrier protein reductase
                       (fabG)"
                       /protein_id="AAD07627.1"
                       /db_xref="GI:2313678"
                       /translation="MQFTGKNVLITGASKGIGAEIAKTLASMGLKVWINYRSNAEVAD
                       ALKNELEEKGYKAAVIKFDAASESDFIEAIQTIVQSDGGLSYLVNNAGVVRDKLAIKM
                       KTEDFHHVIDNNLTSAFIGCREALKVMSKSRFGSVVNVASIIGERGNMGQTNYSASKG
                       GMIAMSKSFAYEGALRNIRFNSVTPGFIETDMNANLKDELKADYVKNIPLNRLGSAKE
                       VAEAVAFLLSDHSSYITGETLKVNGGLYM"
     gene              complement(4732. .4944)
                       /gene="HP0562"
     CDS               complement(4732. .4944)
                       /gene="HP0562"
                       /note="similar to SP:P49225 PID:710340 percent identity:
                       42.37; identified by sequence similarity; putative"
                       /codon_start=1
                       /transl_table=11
                       /product="ribosomal protein S21 (rps21)"
                       /protein_id="AAD07628.1"
                       /db_xref="GI:2313679"
                       /translation="MPGIKVREGDAFDEAYRRFKKQTDRNLVVTECRARRFFESKTEK
                       RKKQKISAKKKVLKRLYMLRRYESRL"
     gene              complement(5045. .6295)
                       /gene="HP0563"
     CDS               complement(5045. .6295)
                       /gene="HP0563"
                       /note="hypothetical protein; identified by GeneMark;
                       putative"
                       /codon_start=1
                       /transl_table=11
                       /product="H. pylori predicted coding region HP0563"
                       /protein_id="AAD07632.1"
                       /db_xref="GI:2313683"
                       /translation="MAINTFLKHSFLVCLLAVNSYAFDWNIFKYNLGFNMFIMDHEGS
                       TPYWVNTNTNLKTRLTPNFGIQFYTRGVEQSLTVGAYFFQNFHNYSTNFPYRWGPTMY
                       YKARGKRFTFYGGIFPRKNLLGRYGLNIFAPYYWFIDPNARGFLLQFQNHYSPSKPYY
                       GHAEFMLDWFGGNCYNTCKFGRNPYGNAMDRFQMNGSVAYNFFKDLLGIGGYFVLPHN
                       EDKYLLNGADGMQFNEKKAIDNSAIYLLDRLYYNAYISTSLLDIAPFMEKLSAKFGMV
                       SEASRLRNREKEVPFINSVGGQFDVEIQYKGFGIHNLFFFAKTPEMPFYNQYQYVEMY
                       CTPSYCPTPIYRGVPFFQAHMYNREFDFYINWKNDFASVRINFVLNAMRGGFDRSLPWS
                       ESYQVYMTVAFDPYNLINKIARKK"
```



```
     gene              6543. .6713
                       /gene="HP0564"
     CDS               6543. .6713
                       /gene="HP0564"
                       /note="hypothetical protein; identified by GeneMark;
                       putative"
                       /codon_start=1
                       /transl_table=11
                       /product="H. pylori predicted coding region HP0564"
                       /protein_id="AAD07634.1"
                       /db_xref="GI:2313685"
                       /translation="MTFYLSKDEHDVLRRLADEEVESVNSFVKRHILKTIIYKKGTNQ
                       DSSINCDSSSRL"
     gene              complement(6736. .7383)
                       /gene="HP0565"
     CDS               complement(6736. .7383)
                       /gene="HP0565"
                       /note="hypothetical protein; identified by GeneMark;
                       putative"
                       /codon_start=1
                       /transl_table=11
                       /product="H. pylori predicted coding region HP0565"
                       /protein_id="AAD07629.1"
                       /db_xref="GI:2313680"

  Query Match             0.9%;  Score 55;  DB 1;  Length 10482;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-15;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    4332 acgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaaaagact 4386
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3267 ACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAAAAGACT 3321


RESULT 200
AE001535
LOCUS       AE001535    11148 bp    DNA             BCT       20-JAN-1999
DEFINITION  Helicobacter pylori, strain J99 section 96 of 132 of the complete
            genome.
ACCESSION   AE001535 AE001439
VERSION     AE001535.1  GI:4155664
KEYWORDS    .
SOURCE      Helicobacter pylori J99.
  ORGANISM  Helicobacter pylori J99
            Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
            Helicobacter.
REFERENCE   1  (bases 1 to 11148)
  AUTHORS   Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C.,
            Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G.,
            Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C.,
            Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E.,
```

```
            Vovis,G.F. and Trust,T.J.
  TITLE     Genomic-sequence comparison of two unrelated isolates of the human
            gastric pathogen Helicobacter pylori
  JOURNAL   Nature 397 (6715), 176-180 (1999)
  MEDLINE   99120557
  REMARK    Erratum:[(published erratum appears in Nature 1999 Feb
            25;397(6721):719)]
REFERENCE   2  (bases 1 to 11148)
  AUTHORS   King,B.L., Alm,R.A. and Trust,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney
            Street, Cambridge, MA 02139, USA
  COMMENT   Address all correspondence to:  hp@arcb.us.astra.com  or  Richard
            A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
            MA,  02139.  Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,
            Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
            Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
            Corporation, 100 Beaver Street, Waltham, MA, 02453.  Qin Jiang and
            Diane E. Taylor are with the University of Alberta Department of
            Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G
            2H7 and the Canadian Bacterial Diseases Network.  All other authors
            are with Astra Research Center Boston, 128 Sidney Street,
            Cambridge, MA, 02139.  Putative identifications, sequence
            alignments, and name and sequence search capability are available
            at ARCB's World Wide Web site.  (URL:
            http://www.astra-boston.com/hpylori).
FEATURES            Location/Qualifiers
     source         1. .11148
                    /organism="Helicobacter pylori J99"
                    /strain="J99"
                    /db_xref="taxon:85963"
     gene           complement(1460. .2317)
                    /gene="jhp1073"
     CDS            complement(1460. .2317)
                    /gene="jhp1073"
                    /note="similar to H. pylori 26695 gene HP1145 and HP1146"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06653.1"
                    /db_xref="GI:4155667"
                    /translation="MPRSPCYTQPLCVAKFLFFRKNMGKTETTIFVDWENLLADLKAI
                    QETDERLKESNFNFNNPEQLLALIRSFLEPEEELKRIYFYVSEPFTEVEPRIKSNKKE
                    ELEEYKEKNPKEYEEKVNKSGIIQSFNHAIAQQNQVKLRVGRVRFMFKDMPKDQRVHG
                    CLEAELILPHLELRQKQIDALAHDITKLXCTKPLGCVVLFSKDTDFVPVLEAAWEKG
                    FEVFIANIQESPNSVPSDLKKSCNVRERNVAEIVDNLPKNQYTPKKKNFSPNEPFNNP
                    FKDQLFKKN"
     gene           complement(3170. .3526)
                    /gene="rplS"
                    /note="jhp1074"
     CDS            complement(3170. .3526)
```



```
                    /gene="rplS"
                    /note="similar to H. pylori 26695 gene HP1147"
                    /codon_start=1
                    /transl_table=11
                    /product="50S RIBOSOMAL PROTEIN L19."
                    /protein_id="AAD06654.1"
                    /db_xref="GI:4155668"
                    /translation="MKNRYIQQFEDAQLKDKSMPTFKAGDTLRLGITIKEGEKTRTQY
                    FEGVCIAIRGNGVDKTFCVRKIGANNIGVEKIFPFYSESLASVEVLRVGRVRRAKLYY
                    LRDRRGKAARIKEVRH"
     gene           complement(3548. .4237)
                    /gene="trmD"
                    /note="jhp1075"
     CDS            complement(3548. .4237)
                    /gene="trmD"
                    /note="similar to H. pylori 26695 gene HP1148"
                    /codon_start=1
                    /transl_table=11
                    /product="TRNA (GUANINE-N1)-METHYLTRANSFERASE"
                    /protein_id="AAD06655.1"
                    /db_xref="GI:4155669"
                    /translation="MKFSVLTLFPQLILPYFEDSILKRALEKNLFELEVLNLRDFSAN
                    KYQKADHTLIGGGAGQILDPEMIENALHSVKNPKHTIFLSAVGKPFKQIDAMRLAQKK
                    HVVLVCGRYEGFDERSIELGADEVFCIGDFILTGGELGALCLIDSIARHIQGVLGNAQ
                    SLENESFENNYLETPNFANAVFKSKEINNIPAPLEYSKGNHAKIKQLKLDLSKLRTKF
                    YRLDLFKQHKS"
     gene           complement(4238. .4783)
                    /gene="jhp1076"
     CDS            complement(4238. .4783)
                    /gene="jhp1076"
                    /note="similar to H. pylori 26695 gene HP1149"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06656.1"
                    /db_xref="GI:4155670"
                    /translation="MLLVGRIGKSVGLNGGLKLHLESDFPECLKKGVKVSVAPINAFS
                    CASSFKDYVIHSYEHAKNLLPLETIRTPEKAKELTNLGLFMSEAESKKLCVLKEGEFF
                    YCDLIGLSVVEGNEILGKVIEIQRISQTDYFMVETTRSLVEKGLAKIFLIPYRDFYIQ
                    EILLQDKKITTHNAKTLLENS"
     gene           complement(4793. .5146)
                    /gene="jhp1077"
     CDS            complement(4793. .5146)
                    /gene="jhp1077"
                    /note="similar to H. pylori 26695 gene HP1150"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06657.1"
                    /db_xref="GI:4155671"
                    /translation="MRKSDPLNTPFSQAECKDYSYCVATFLEKYLKKVVSFPQALSVE
```



```
                    YTLLEDKVKQITIYTHPSDMGHVIGKEGKMVSAIKAFVSGVKAKDGFSYKIVVFASND
                    KNGDKNPHALGDQTP"
     gene           complement(5163. .5393)
                    /gene="rpsP"
                    /note="jhp1078"
     CDS            complement(5163. .5393)
                    /gene="rpsP"
                    /note="similar to H. pylori 26695 gene HP1151"
                    /codon_start=1
                    /transl_table=11
                    /product="30S RIBOSOMAL PROTEIN S16"
                    /protein_id="AAD06658.1"
                    /db_xref="GI:4155672"
                    /translation="MTVIRLTRIGRKKKPFYRVVVTDSRKRRDGGWIESIGYYNPLSE
                    PKDVRIDKERLNYWKGVGAKMSERVEKLSQKA"
     gene           complement(5467. .6813)
                    /gene="ffh"
                    /note="jhp1079"
     CDS            complement(5467. .6813)
                    /gene="ffh"
                    /note="similar to H. pylori 26695 gene HP1152"
                    /codon_start=1
                    /transl_table=11
                    /product="putative SIGNAL RECOGNITION PARTICLE PROTEIN"
                    /protein_id="AAD06659.1"
                    /db_xref="GI:4155673"
                    /translation="MFQALSDGFKNALNKIRPQDDEKALDRALDELKKTLLKNDVHHK
                    VARELLKKVESQTKAGGIGKQQFLDALEKSLLEILSAKGSSGPTFAQTPPTVVLMAGL
                    QGSGKTTTTAKLAHYLKTKNKKVLLCACDLQRLAAVEQLKVLGEQVGVEVFHEENKSV
                    KEIANNALKRAKEAQFDVLIVDSAGRLAIDKELMQELKEVKEVLNPHEVLYVADALSG
                    QDGVKSANTFNEEIGVSGVVLSKFDSDSKGGIALGITYQLGLPLRFIGSGEKIPDLDV
                    FMPERIVGRLMGAGDIISLAEKTASVLNPNEAKDLSKKLKKGQFTFNDFLNQIEKVKK
                    LGSMSSLISMIPGLGNMASALKDTDLESSLEVKKIKAMVNSMTKKERENPEILNGSRR
                    KRIALGSGLEVSEINRIIKRFDQASKMAKRLTNKKGISDLMNLMSQAKNQTPPKMR"
     gene           complement(6828. .9446)
                    /gene="valS"
                    /note="jhp1080"
     CDS            complement(6828. .9446)
                    /gene="valS"
                    /note="similar to H. pylori 26695 gene HP1153"
                    /codon_start=1
                    /transl_table=11
                    /product="VALYL-TRNA SYNTHETASE"
                    /protein_id="AAD06660.1"
                    /db_xref="GI:4155674"
                    /translation="MKQEPTTYQPEEIEKKIYEICSHRGYFEINGNEKIQEKGKRFCL
                    MMPPPNVTGILHIGHALTLSLQDILVRYKRMDGYKTLYQPGLDHAGIATQNVVEKQLL
                    SQGVKKEDLGREKFIQKVWEWKEKSGGAILEQMKRLGVSTAFSRTRFTMDKGLQRAVK
                    LAFLKWYEKGLIVQDNYMVNWCTKDGALSDIEVEYEERKGALYYIRYYLENQKDYLVV
                    ATTRPETLFGDSAIMVNPNDERYKHLVGQQVILPLINRTIPIIADAHVEMGFGTGCVK
                    VTPGHDFNDYEVGKRHHLETIKIFDEKGILNAHCGEFENLERLEARDKVVAALKENAL
```



```
                    LEKIEEHVHQVGHCYRCHNVVEPYVSKQWFVKPEIAQSSIEKIQQGLARFYPSNWINN
                    YNAWMRELRPWCISRQLFWGHQIPVFTCENNHQFVSLDTPLSCPTCKSEKLEQDKDVL
                    DTWFSSGLWAFSTLGWGQEKSDLFNESDLKDFYPNTTLITGFDILFFWVARMLFCSES
                    LLGELPFKDIIYLHALVRDEKGEKMSKSKGNVIDPLEMIEKYGADSLRFTLANLCATGR
                    DIKLSTTHLENNKNFANKLFNAASYLKLKQESFKDKERLNEIQTALGRYAKSRLNLVF
                    KEVRNALDNYFRNDATTLSFRFLWGEFCDWFIEFSKVENEAIDELGSVLKEALKLLHP
                    FMPFISESLYHKLSNTELENAHSIMVMPYPKEIAQDEKLEHEFEVIKDCIVSLRRLKI
                    MLETPPIVLKEASVGLREKIENTERLQNYAQKLAKLEKVSVITYKPLKSVSDVGEFCQ
                    TYADLENLDLSPLIARLKKQLEKLEKEKLKLNLHNENFVKNAPKSVLEKARESLKTLL
                    EKEGKIQQELDLLEQP"
     gene           9584. .9991
                    /gene="jhp1081"
     CDS            9584. .9991
                    /gene="jhp1081"
                    /note="similar to H. pylori 26695 gene HP1154"
                    /codon_start=1
                    /transl_table=11
                    /product="putative"
                    /protein_id="AAD06651.1"
                    /db_xref="GI:4155665"
                    /translation="MNYFLKAPILGFEHINEVRLEKIDSLFSRLMGQTNSPMALDMVL
                    VNPYCLREYSFVIPKYIELLLELDSHSKVEVYCVVVLQKNLEDSMVNFLAPLVFNSKN
                    GFGAQVALSMMDYPDFGFRDPLKSFVVKERERA"
     gene           10003. .11064
                    /gene="murG"
                    /note="jhp1082"
     CDS            10003. .11064
                    /gene="murG"
                    /note="similar to H. pylori 26695 gene HP1155"
                    /codon_start=1
                    /transl_table=11
                    /product="UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-
```

```
  Query Match             0.9%;  Score 55;  DB 1;  Length 11148;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-15;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4074 tttgggatgaatgctaccaattcatggtatcatatcccatacattcgtatctag 4128
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2659 TTTGGGATGAATGCTACCAATTCATGGTATCATATCCCCATACATTCGTATCTAG 2713


Search completed: March 26, 2001, 22:17:03
Job time: 34281 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 14:45:04 ; Search time 371.66 Seconds
(without alignments)
5988.811 Million cell updates/sec

Title: US-09-410-835-4
Perfect score: 5925
Sequence: 1 ctccattttaagcaactcca.....ttttagaatagacaaagctt 5925

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 480022 seqs, 187831343 residues

Word size : 15

Total number of hits satisfying chosen parameters: 8150

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N_Geneseq_36:*
1: /cgn1_8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn1_8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn1_8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn1_8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn1_8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn1_8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn1_8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn1_8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn1_8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn1_8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn1_8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn1_8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn1_8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn1_8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn1_8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn1_8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn1_8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn1_8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn1_8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn1_8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn1_8/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5925	100.0	5925	14	Q48733	CAI gene. Helicob
2	5925	100.0	19932	17	T46159	CagI locus. Helic
3	3444	58.1	4149	19	V62461	Helicobacter pylor
4	277	4.7	449	19	V73601	H. pylori cagA DNA
5	277	4.7	449	19	V73602	H. pylori cagA DNA
6	272	4.6	449	19	V73610	H. pylori cagA DNA
7	235	4.0	3648	15	Q64580	120-128 kilodalton
8	235	4.0	3648	17	T14052	Helicobacter pylor
9	235	4.0	4821	15	Q64581	Tag A antigen of H
10	235	4.0	4821	16	Q86728	H. pylori tagA ant
11	235	4.0	4821	17	T14051	Helicobacter pylor
12	210	3.5	449	19	V73600	H. pylori cagA DNA
13	194	3.3	449	19	V73606	H. pylori cagA DNA
14	179	3.0	449	19	V73609	H. pylori cagA DNA
15	179	3.0	449	19	V73611	H. pylori cagA DNA
16	175	3.0	449	19	V73604	H. pylori cagA DNA
17	166	2.8	3534	18	T67410	H. pylori cytoplas
18	166	2.8	3549	18	T68137	H. pylori cytoplas
19	156	2.6	449	19	V73605	H. pylori cagA DNA
20	136	2.3	449	19	V73608	H. pylori cagA DNA
21	131	2.2	449	19	V73612	H. pylori cagA DNA
22	125	2.1	449	19	V73607	H. pylori cagA DNA
23	121	2.0	449	19	V73613	H. pylori cagA DNA
24	119	2.0	449	19	V73603	H. pylori cagA DNA
c 25	110	1.9	285	20	V90655	Nucleotide sequenc
26	110	1.9	2705	20	V90548	Nucleotide sequenc
27	110	1.9	2705	20	V90861	Nucleotide sequenc
c 28	79	1.3	386	19	X13977	H. pylori GHPO 178
c 29	67	1.1	564	18	T67571	H. pylori cytoplas
c 30	67	1.1	657	18	T68133	H. pylori cytoplas
31	61	1.0	239	20	V90662	Nucleotide sequenc
32	61	1.0	449	19	V73614	H. pylori cagA DNA
33	56	0.9	464	19	V73618	H. pylori cagA DNA
34	50	0.8	464	19	V73621	H. pylori cagA DNA
35	49	0.8	464	19	V73616	H. pylori cagA DNA
36	49	0.8	464	19	V73617	H. pylori cagA DNA
37	49	0.8	464	19	V73620	H. pylori cagA DNA
38	40	0.7	464	19	V73619	H. pylori cagA DNA
39	38	0.6	464	19	V73622	H. pylori cagA DNA
40	34	0.6	464	19	V73615	H. pylori cagA DNA
41	31	0.5	37	19	V45503	Helicobacter pylor

c 42	31	0.5	990	19	X14282	H. pylori GHPO 174
43	27	0.5	36	21	Z50141	CAG 1 PCR primer f
44	26	0.4	35	21	Z50143	CAG 3 PCR primer f
c 45	25	0.4	28	19	V65776	Helicobacter pylor
c 46	25	0.4	34	21	Z50142	CAG 2 PCR primer f
47	24	0.4	24	19	V73519	H. pylori cag A pr
c 48	23	0.4	30	19	V45504	Helicobacter pylor
49	23	0.4	2498	19	X14522	H. pylori GHPO 130
c 50	22	0.4	22	19	V73520	H. pylori cag A pr
51	22	0.4	267	20	V90656	Nucleotide sequenc
c 52	21	0.4	21	19	V73527	H. pylori cag A pr
53	21	0.4	1603	19	V32400	Homo sapiens spryl
54	21	0.4	2030	21	A14073	Human SPROUTY-1 pa
55	21	0.4	2138	21	A14073	Human SPROUTY-1 pa
56	20	0.3	20	18	V25297	Primer F2 for H.py
57	20	0.3	20	19	V73508	H. pylori cag A pr
c 58	20	0.3	20	19	V73534	H. pylori cag A pr
c 59	20	0.3	20	19	V50362	Antisense oligomer
60	20	0.3	21	18	V25299	Primer F3 for H.py
61	20	0.3	284	14	Q38651	Truncated Rs-APP2
62	20	0.3	284	15	Q70129	Antimicrobial Rs-A
c 63	20	0.3	302	21	A06488	Human immunogenic
64	20	0.3	610	20	X61423	DNA encoding a hum
65	20	0.3	619	21	A26393	Human secreted pro
66	20	0.3	1118	21	A09041	Human TANGO 195 pa
67	20	0.3	1440	18	V75036	Staphylococcus aur
68	20	0.3	1537	18	T72662	Human smooth muscl
69	20	0.3	1537	20	X90357	Human smooth muscl
70	20	0.3	1969	19	V34250	Human secreted pro
71	20	0.3	2187	19	V38661	Mus musculus SOCS3
72	20	0.3	2201	20	X90601	Histidine kinase e
73	20	0.3	2309	15	Q73786	Partial PCR fragme
74	20	0.3	2378	19	V34188	Human secreted pro
c 75	20	0.3	2529	13	Q20576	Transforming Growt
c 76	20	0.3	2530	11	Q06845	Sequence encpding
77	20	0.3	2734	14	Q39221	Encodes human epid
c 78	20	0.3	2734	14	Q47376	hETG mRNA. Homo s
79	20	0.3	2894	21	A09042	Full-length human
c 80	20	0.3	14796	19	V27941	Survivin gene. Ho
c 81	20	0.3	25871	21	A09888	Human genomic OCTN
82	19	0.3	19	19	V73526	H. pylori cag A pr
83	19	0.3	28	18	V25348	Primer for H.pylor
c 84	19	0.3	29	18	V25347	Primer for H.pylor
c 85	19	0.3	35	16	Q87899	Normalised library
86	19	0.3	625	20	X97725	Extended human sec
87	19	0.3	760	20	Z52996	Human prostate tum
c 88	19	0.3	801	21	Z39022	Mouse novel cDNA c
c 89	19	0.3	879	18	V24653	H. pylori ORF 05ae
c 90	19	0.3	991	20	Z00843	Human secreted pro
91	19	0.3	1015	20	X00713	Human secreted pro
92	19	0.3	1046	19	V82785	Clone ek591_1 isol

93	19	0.3	1213	19	V53421	DNA encoding a Sta
94	19	0.3	1580	20	X00661	Human secreted pro
c 95	19	0.3	1786	19	X14271	H. pylori GHPO 162
96	19	0.3	2042	20	X00621	Human secreted pro
97	19	0.3	2457	21	A26310	Human secreted pro
c 98	19	0.3	2547	19	V61484	Human secreted pro
c 99	19	0.3	2883	18	T58898	C-Delta-1 gene (al
100	19	0.3	4089	17	T16857	Bovine endothelial
c 101	19	0.3	4748	20	X76320	Schizosaccharomyce
c 102	19	0.3	4748	20	X57223	S. pombe invertase
c 103	19	0.3	5790	20	X24464	Bovine myostatin g
104	19	0.3	7179	20	Z21006	Salmonella dublin
c 105	19	0.3	7286	20	X57224	WO 9923223 Seq ID
c 106	19	0.3	7938	20	X57235	WO 9923223 Seq ID
c 107	19	0.3	14000	21	A09020	A. thaliana ATPAC
108	19	0.3	15462	19	V70401	Human parainfluenz
109	19	0.3	15512	19	V59470	Human ryanodin rec
110	19	0.3	25464	19	V57274	Human flavin-conta
c 111	19	0.3	25580	20	X12956	Enterococcus faeca
c 112	19	0.3	58407	19	V21210	Methanococcus jann
113	19	0.3	68940	20	X57351	Human chromosome 6
c 114	19	0.3	162450	21	Z06967	Retinoblastoma bin
c 115	18	0.3	18	19	V73529	H. pylori cag A pr
116	18	0.3	18	19	V65775	Helicobacter pylor
c 117	18	0.3	21	16	Q75761	Reverse transcript
c 118	18	0.3	40	18	T61445	Nucleotide anchore
c 119	18	0.3	40	20	Z25103	Human ischaemic he
c 120	18	0.3	40	20	X36067	Primer used to syn
c 121	18	0.3	40	21	Z51197	Primer RP6.0, for
c 122	18	0.3	40	21	Z59453	PCR primer RP6.0 u
c 123	18	0.3	40	21	Z38052	Jurkat cell cDNA
c 124	18	0.3	41	20	X34852	Primer for adapter
c 125	18	0.3	41	20	X23456	Human neutrophil c
126	18	0.3	43	19	V48090	Oligonucleotide 43
127	18	0.3	65	21	A08140	MOMP containing fu
c 128	18	0.3	65	21	A08141	MOMP containing fu
129	18	0.3	115	20	V90176	EST clone DP776.
130	18	0.3	226	14	Q37967	Sequence encoding
131	18	0.3	226	19	V60636	5'RACE product fro
132	18	0.3	256	20	V88786	EST clone HO332.
c 133	18	0.3	281	20	V89295	EST clone CH224.
134	18	0.3	355	14	Q39730	Expressed Sequence
135	18	0.3	355	14	Q59142	Human brain Expres
136	18	0.3	360	12	Q10253	Synthetic version
137	18	0.3	367	20	Z24905	Human secreted pro
c 138	18	0.3	369	21	Z54442	Neisseria meningit
c 139	18	0.3	425	18	X30899	Streptococcus pneu
140	18	0.3	471	20	X97675	Extended human sec
c 141	18	0.3	491	20	V86280	EST clone AJ143.
142	18	0.3	507	19	V09106	3' nucleotide sequ
143	18	0.3	507	20	X77537	Human secreted pro

144	18	0.3	507	21	Z59479	Human secreted pro
c 145	18	0.3	516	21	Z54438	Neisseria gonorrhe
c 146	18	0.3	516	21	Z54440	Neisseria gonorrhe
c 147	18	0.3	518	19	T98751	DNA encoding a S.
c 148	18	0.3	522	21	Z54441	Neisseria meningit
149	18	0.3	542	20	X97614	Extended human sec
150	18	0.3	545	20	X20418	Human secreted pro
151	18	0.3	566	18	V02880	Human HMGI-C aberr
152	18	0.3	613	19	V03044	Rat EX clone nucle
c 153	18	0.3	634	20	V86929	EST clone BK299.
154	18	0.3	637	20	X22221	Human secreted pro
c 155	18	0.3	664	19	Z96205	S. pneumoniae deri
156	18	0.3	675	20	X97861	Human secreted pro
c 157	18	0.3	676	20	Z16920	Human gene express
158	18	0.3	680	21	A26408	Human secreted pro
159	18	0.3	693	18	V24706	H. pylori ORF 14ee
160	18	0.3	718	21	Z56708	Human transmembran
161	18	0.3	741	20	Z17597	Human gene express
162	18	0.3	742	20	Z15143	Human gene express
163	18	0.3	756	21	A37024	Human PRO1018 (UNQ
164	18	0.3	774	18	V25011	H. pylori inner me
c 165	18	0.3	802	20	Z52677	Human prostate tum
166	18	0.3	810	19	X14350	H. pylori GHPO 292
c 167	18	0.3	811	19	V11433	Human secreted pro
168	18	0.3	840	20	X30394	DNA encoding a hum
169	18	0.3	855	20	X97814	Human secreted pro
170	18	0.3	861	20	Z07553	Maize glutathione-
171	18	0.3	861	21	Z94793	Maize class III gl
172	18	0.3	882	18	T72173	Alzheimer's diseas
173	18	0.3	899	12	Q14451	Rat prolactin gene
174	18	0.3	924	13	Q31989	Guinea pig MBP-2 g
175	18	0.3	940	16	Q85268	Spider venom FIL-7
176	18	0.3	990	21	Z97102	Human secreted pro
177	18	0.3	1002	20	X58797	Human leukaemia-as
178	18	0.3	1047	21	A08124	C. trachomatis MOM
179	18	0.3	1051	11	Q04287	Sequence encoding
180	18	0.3	1057	11	Q04285	cDNA clone of sequ
181	18	0.3	1080	20	Z00410	Human secreted pro
c 182	18	0.3	1083	20	X08684	Novel nucleotide s
183	18	0.3	1100	20	X60109	DNA encoding a for
184	18	0.3	1134	20	X00711	Human secreted pro
185	18	0.3	1147	20	X00658	Human secreted pro
186	18	0.3	1197	20	X30152	Human secreted pro
187	18	0.3	1204	21	A37072	Human PRO1279 (UNQ
188	18	0.3	1223	20	Z33523	Human prostate can
c 189	18	0.3	1230	18	T67029	Salmonella secrete
190	18	0.3	1240	20	Z41296	Human normal ovari
191	18	0.3	1250	19	V26294	Phosphate starvati
c 192	18	0.3	1254	18	T43482	ATM gene exons 48-
193	18	0.3	1265	20	Z33591	Human breast tumou
194	18	0.3	1290	12	Q15023	Rat prolactin codi

195	18	0.3	1303	20	Z42221	Human normal bladd
196	18	0.3	1314	21	Z30222	cDNA encoding a hu
197	18	0.3	1350	21	A26298	Human secreted pro
c 198	18	0.3	1357	12	Q11505	Yeast promoter for
199	18	0.3	1358	19	V38741	cDNA nfSPI2-1358 e
c 200	18	0.3	1358	19	V38742	Complementary stra
201	18	0.3	1373	12	Q15501	Interleukin-1 beta
202	18	0.3	1373	14	Q38669	Interleukin 1-beta
203	18	0.3	1373	15	Q54402	ICE gene. Homo sa
204	18	0.3	1373	17	T38197	Human interleukin-
205	18	0.3	1379	5	N40222	Human IgD. Homo s
206	18	0.3	1414	19	V38745	cDNA nfSPI4-1414 e
c 207	18	0.3	1414	19	V38746	Complementary stra
208	18	0.3	1454	19	V38749	cDNA nfSPI6-1454 e
c 209	18	0.3	1454	19	V38750	Complementary stra
210	18	0.3	1464	20	Z41351	Human normal uteru
c 211	18	0.3	1483	12	Q11506	Yeast promoter for
c 212	18	0.3	1483	12	Q14417	Yeast glycerol-3-p
213	18	0.3	1492	19	V38747	cDNA nfSPI5-1492 e
c 214	18	0.3	1492	19	V38748	Complementary stra
215	18	0.3	1494	18	T59907	Yeast transcriptio
216	18	0.3	1508	17	T37312	Aromatic acyl tran
217	18	0.3	1539	17	T39751	Human mucosal addr
218	18	0.3	1539	18	X35280	cDNA encoding huma
219	18	0.3	1560	13	Q29613	Encodes carp TLAIa
220	18	0.3	1560	20	X27334	Human secreted pro
221	18	0.3	1573	15	Q68222	Neuroendocrine tum
c 222	18	0.3	1584	19	V38740	Complementary stra
223	18	0.3	1584	19	V38739	cDNA nfSPI1-1584 e
224	18	0.3	1619	21	Z29079	A. palaestina cDNA
225	18	0.3	1659	12	Q14265	Interleukin-1 b pr
226	18	0.3	1694	17	T15232	Tumor necrosis fac
227	18	0.3	1694	18	T94635	TNF-R1-DD ligand p
228	18	0.3	1715	20	Z24887	Human secreted pro
229	18	0.3	1715	20	X85019	Human secreted pro
230	18	0.3	1822	12	Q11803	Sequence encoding
231	18	0.3	1822	13	Q24518	Human osteogenic p
232	18	0.3	1822	13	Q28735	Human osteogenic p
233	18	0.3	1822	14	Q53141	Sequence encoding
234	18	0.3	1822	14	Q38944	Human osteogenic p
235	18	0.3	1822	14	Q38733	Human OP-1. Homo
236	18	0.3	1822	14	Q38857	Morphogen hOP1 cod
237	18	0.3	1822	15	Q54755	Osteogenic protein
238	18	0.3	1822	15	Q56198	hOP1 cDNA. Homo s
239	18	0.3	1822	15	Q57916	Human osteogenic p
240	18	0.3	1822	15	Q67311	Human OP-1. Homo
241	18	0.3	1822	15	Q45116	Human OP-1. Homo
242	18	0.3	1822	15	Q65391	Osteogenic protein
243	18	0.3	1822	15	Q45162	Human OP-1. Homo
244	18	0.3	1822	15	Q71425	Osteogenic protein
245	18	0.3	1822	15	Q72703	hOP1 human osteoge

246	18	0.3	1822	16	T06038	Human osteogenic p
247	18	0.3	1822	17	T33441	Human osteogenic p
248	18	0.3	1822	17	T06771	Human osteogenic p
249	18	0.3	1822	17	T02597	hOP-1 cDNA. Homo
250	18	0.3	1822	18	V10345	Human OP-1 cDNA.
251	18	0.3	1822	18	T93810	Human osteogenic p
252	18	0.3	1822	18	T97877	cDNA encoding huma
253	18	0.3	1822	18	T87879	Human osteogenic p
254	18	0.3	1822	18	T73207	Human hippocampal
255	18	0.3	1822	19	V65199	Human osteogenic p
256	18	0.3	1822	19	V60227	Nucleotide sequenc
257	18	0.3	1822	19	V41572	Morphogenically ac
258	18	0.3	1822	19	V32583	Human osteogenic p
259	18	0.3	1822	19	V19533	Human osteogenic p
260	18	0.3	1822	19	V15204	Human osteogenic p
261	18	0.3	1822	20	Z30824	Human osteogenic p
262	18	0.3	1822	20	Z27574	Human osteogenic p
263	18	0.3	1822	20	X80559	Human osteogenic p
264	18	0.3	1822	20	X83980	Human osteogenic p
265	18	0.3	1822	20	X79684	Human OP-1 coding
266	18	0.3	1822	20	X56414	Human osteogenic p
267	18	0.3	1822	20	V80735	Human osteogenic p
268	18	0.3	1822	20	X00229	Human osteogenic p
269	18	0.3	1822	20	V80656	Human osteogenic p
270	18	0.3	1822	21	A09361	Human osteogenic p
271	18	0.3	1822	21	A09390	Morphogenic protei
272	18	0.3	1822	21	Z95567	Human osteogenic p
273	18	0.3	1822	21	Z39477	Human osteogenic p
274	18	0.3	1823	15	Q56231	hOP1 cDNA. Homo s
275	18	0.3	1838	19	V38743	cDNA nfSPI3-1838 e
c 276	18	0.3	1838	19	V38744	Complementary stra
277	18	0.3	1855	20	X87688	Human tyrosylprote
278	18	0.3	1855	21	Z99699	cDNA encoding a hu
279	18	0.3	1872	16	Q90623	Murine osteogenic
280	18	0.3	1872	17	T33442	Mouse osteogenic p
281	18	0.3	1873	13	Q24517	Murine osteogenic
282	18	0.3	1873	13	Q28736	Murine osteogenic
283	18	0.3	1873	14	Q53153	Sequence encoding
284	18	0.3	1873	14	Q38945	Mouse osteogenic p
285	18	0.3	1873	14	Q38734	Murine pro-OP-1.
286	18	0.3	1873	14	Q38858	Morphogen mOP1 cod
287	18	0.3	1873	15	Q56199	mOP1 cDNA. Mus mu
288	18	0.3	1873	15	Q56232	mOP1 cDNA. Mus mu
289	18	0.3	1873	15	Q58051	Mouse osteogenic p
290	18	0.3	1873	15	Q67312	Murine OP-1. Mus
291	18	0.3	1873	15	Q45117	Murine OP-1. Mus
292	18	0.3	1873	15	Q65392	Osteogenic protein
293	18	0.3	1873	15	Q45163	Murine OP-1. Mus
294	18	0.3	1873	15	Q72704	mOP1-PP prepro for
295	18	0.3	1873	17	T02598	mOP-1 cDNA. Mus m
296	18	0.3	1873	18	V10346	Mouse OP-1 cDNA.

297	18	0.3	1873	18	T97879	cDNA encoding mous
298	18	0.3	1873	19	V32584	Mouse osteogenic p
299	18	0.3	1873	19	V19534	Mouse osteogenic p
300	18	0.3	1873	19	V15216	Mouse osteogenic p
301	18	0.3	1873	20	Z27576	Mouse osteogenic p
302	18	0.3	1873	20	X00241	Murine osteogenic
303	18	0.3	1873	21	Z39478	Mouse osteogenic p
304	18	0.3	1882	16	Q90522	Oxalyl-CoA decarbo
305	18	0.3	1882	18	T75517	cDNA encoding huma
306	18	0.3	1905	17	T33552	Glutathione synthe
307	18	0.3	1906	11	Q04531	Human prolidase ge
308	18	0.3	1975	14	Q36031	Chicken nov coding
309	18	0.3	1982	19	V10366	Wheat lov affinity
310	18	0.3	2083	21	Z91091	Human putative SST
c 311	18	0.3	2090	21	Z97041	Human secreted pro
312	18	0.3	2148	20	Z24650	Human lung tumor a
313	18	0.3	2164	18	T58559	Human breast cance
314	18	0.3	2164	20	X04731	Full cDNA sequence
315	18	0.3	2164	21	A07221	Full length gp30/a
316	18	0.3	2199	18	T58559	Human breast cance
317	18	0.3	2199	20	X87701	Herugelin-beta-1 c
318	18	0.3	2199	20	X04732	Full cDNA sequence
319	18	0.3	2199	21	A07222	Full length gp30/b
320	18	0.3	2215	21	Z43884	C. vicina LSP-2 cD
321	18	0.3	2226	20	X87700	Herugelin-alpha cD
322	18	0.3	2313	20	Z07197	Human lung tumour
323	18	0.3	2333	21	Z50745	Human sensory tran
324	18	0.3	2383	21	A35134	Human adenosine re
325	18	0.3	2524	17	T18673	Human trophinin cD
326	18	0.3	2547	19	V61484	Human secreted pro
327	18	0.3	2571	20	X13144	Enterococcus faeca
328	18	0.3	2582	18	T87428	Clone J422. Homo
329	18	0.3	2595	21	Z61177	Genetic sequence o
330	18	0.3	2599	20	X33957	Human HCMV repress
331	18	0.3	2599	21	A34742	Human adenosine re
332	18	0.3	2599	21	A34755	Human adenosine re
333	18	0.3	2599	21	A34764	Human adenosine re
334	18	0.3	2610	10	N93060	cDNA of plasmid pX
335	18	0.3	2774	12	Q14293	Modified Protein C
336	18	0.3	2781	12	Q13239	HSF cDNA sequence.
337	18	0.3	2781	13	Q25712	Sequence of Drosop
338	18	0.3	2790	20	Z07171	Human lung tumour
339	18	0.3	2801	19	V40749	C. felis esterase,
c 340	18	0.3	2801	19	V40750	C. felis esterase,
341	18	0.3	2836	19	V40744	C. felis esterase,
c 342	18	0.3	2836	19	V40745	C. felis esterase,
343	18	0.3	2889	21	Z50903	Human receptor-ass
344	18	0.3	2971	21	Z98043	Human secreted pro
345	18	0.3	2999	19	V54584	Human secretory pr
346	18	0.3	2999	20	Z25604	Human secreted pro
347	18	0.3	3056	20	X26149	Human pan-s/tk-1B

	348	18	0.3	3086	15	Q62811 Riboflavin 2 gene
	349	18	0.3	3189	20	X20284 Borrelia burgdorfe
	350	18	0.3	3349	20	X60577 Human disintegrin
c	351	18	0.3	3441	18	T67035 Salmonella virulen
	352	18	0.3	3442	19	V19967 Rabbit excitatory
	353	18	0.3	3521	14	Q41226 Clone GP3 encoding
	354	18	0.3	3521	14	Q41230 Gene encoding majo
	355	18	0.3	3607	20	X87412 Hepatocellular car
c	356	18	0.3	3708	20	Z77530 Human ovarian tumo
	357	18	0.3	3735	12	Q14452 Rat prolactin gene
	358	18	0.3	3975	9	N81157 Malaria-specific g
	359	18	0.3	3975	13	Q22999 SERP gene. Plasmo
	360	18	0.3	4089	20	Z08807 Human cerebral pro
	361	18	0.3	4304	17	T32960 Mouse developmenta
c	362	18	0.3	4336	18	V74857 Staphylococcus aur
	363	18	0.3	4467	20	X05820 Rat pheromone rece
	364	18	0.3	4487	12	Q14290 Protein C gene fro
	365	18	0.3	5010	19	X24042 Rat DH1 cDNA. Rat
c	366	18	0.3	5154	20	X22554 Human ZGGBP1 cDNA.
	367	18	0.3	6124	11	Q03568 Sequence encoding
c	368	18	0.3	6124	11	Q03568 Sequence encoding
c	369	18	0.3	6157	20	X13255 Enterococcus faeca
	370	18	0.3	7146	19	V38933 Nucleic acid seque
	371	18	0.3	7435	21	Z93313 Sequence encoding
	372	18	0.3	7587	12	Q15133 Zucchini ACC synth
	373	18	0.3	7587	19	V15703 Zucchini ACC synth
	374	18	0.3	7746	21	A07836 Human homologue of
c	375	18	0.3	14024	20	X06749 E. coli O157 antig
c	376	18	0.3	14024	21	Z56386 Escherichia coli f
	377	18	0.3	32367	19	V35620 Human SHOX (short
	378	18	0.3	138169	21	A34791 Human adenosine re
	379	18	0.3	141589	21	A35005 Human adenosine re
	380	18	0.3	141589	21	A35030 Human adenosine re
	381	18	0.3	162450	21	Z86967 Retinoblastoma bin
c	382	18	0.3	235033	19	V57926 Hereditary haemoch
c	383	18	0.3	237326	19	V57903 Hereditary haemoch
	384	18	0.3	1038602	20	Z01425 Complete genome se
c	385	18	0.3	1230025	20	X91990 Nucleotide sequenc
	386	18	0.3	1664976	19	V21209 Methanococcus jann
c	387	17	0.3	20	16	Q75597 Reverse transcript
c	388	17	0.3	21	16	Q75759 Reverse transcript
c	389	17	0.3	21	16	Q75760 Reverse transcript
c	390	17	0.3	21	16	Q75762 Reverse transcript
c	391	17	0.3	26	21	Z59153 Oligonucleotide #1
	392	17	0.3	27	16	Q75368 Oligonucleotide 1
c	393	17	0.3	32	20	V59863 Mutagenic oligonuc
c	394	17	0.3	33	19	V27540 Streptococcus pneu
c	395	17	0.3	41	19	V51079 Maize polymorphic
c	396	17	0.3	54	14	Q39029 HD1(T7) 3' termina
	397	17	0.3	85	17	T39470 Growth regulatory
	398	17	0.3	85	19	V37488 Human growth regul

c	399	17	0.3	98	12	Q11764 Self-complementary
c	400	17	0.3	123	14	Q38819 HBAg N-terminal.
	401	17	0.3	129	20	X40523 Human secreted pro
c	402	17	0.3	150	19	X12026 Human biallelic po
c	403	17	0.3	178	19	V54125 Fragment of membra
c	404	17	0.3	196	21	A00841 Human colon cancer
c	405	17	0.3	204	20	Z33775 Tobacco plant resi
	406	17	0.3	205	16	T19252 Human gene signatu
c	407	17	0.3	214	21	Z43958 S. aureus detectin
c	408	17	0.3	261	14	Q60823 Human brain Expres
c	409	17	0.3	279	20	V87257 EST clone BN422.
	410	17	0.3	300	20	Z14679 Human gene express
	411	17	0.3	302	20	V89492 EST clone CO1232.
	412	17	0.3	315	20	X25174 HIV-1 group O isol
c	413	17	0.3	319	20	V89670 EST clone DA244.
	414	17	0.3	422	17	T33753 Cellubrevin-1 part
c	415	17	0.3	438	20	V88097 EST clone DO949.
	416	17	0.3	460	21	Z33329 Human secreted pro
	417	17	0.3	463	11	Q04055 cDNA encoding nema
c	418	17	0.3	464	14	Q37717 K1ADH4 promoter Sa
	419	17	0.3	495	18	V02892 Human HMGI-C aberr
	420	17	0.3	526	18	V74827 Staphylococcus aur
	421	17	0.3	587	19	V66757 Pathogen response
c	422	17	0.3	592	20	Z42073 Human endometrium
	423	17	0.3	612	20	X37536 Human secreted pro
	424	17	0.3	615	18	V02881 Human HMGI-C aberr
	425	17	0.3	622	20	X37410 Human secreted pro
	426	17	0.3	642	20	X32463 Sequence 1 provide
	427	17	0.3	658	20	X90723 Human yb37_1 secre
c	428	17	0.3	666	14	Q37718 K1ADH4 promoter Sa
	429	17	0.3	674	20	Z42074 Human endometrium
	430	17	0.3	685	21	A26438 Human secreted pro
	431	17	0.3	696	21	A01677 Human colon cancer
	432	17	0.3	712	20	X22215 Human secreted pro
	433	17	0.3	723	19	V27172 Bacillus thuringie
	434	17	0.3	739	14	Q43423 Sequence encoding
	435	17	0.3	743	8	N70812 Sequence encoding
	436	17	0.3	748	12	Q10376 Plasmid pMG4B12 us
	437	17	0.3	754	21	A01699 Human colon cancer
	438	17	0.3	760	18	T71211 Winged bean lysine
c	439	17	0.3	766	20	Z16999 Human gene express
c	440	17	0.3	781	18	V06144 Viral infection ge
c	441	17	0.3	786	20	Z22984 Nucleotide sequenc
c	442	17	0.3	797	19	V36459 caYAE1 DNA sequenc
	443	17	0.3	798	18	T85783 Human Mucin 8 cDNA
	444	17	0.3	798	20	X20443 Human secreted pro
	445	17	0.3	800	20	X90757 Avian Interleukin-
	446	17	0.3	812	20	X15447 cDNA encoding a hu
	447	17	0.3	823	20	X32455 Plasmid pSJ8-2a pa
	448	17	0.3	825	20	Z23125 Corn sucrose trans
	449	17	0.3	832	19	V34302 Human secreted pro

	450	17	0.3	832	20	X13534 Enterococcus faeca
	451	17	0.3	836	20	X04760 DNA encoding a coc
	452	17	0.3	841	19	V04578 Flea serine protea
	453	17	0.3	843	19	X14241 H. pylori GHPO 134
	454	17	0.3	844	19	V34155 Human secreted pro
c	455	17	0.3	847	19	X14465 H. pylori GHPO 100
c	456	17	0.3	851	20	Z22983 Nucleotide sequenc
c	457	17	0.3	851	20	Z22987 Nucleotide sequenc
	458	17	0.3	858	20	X34757 DNA encoding OMP-1
	459	17	0.3	875	19	V58364 Coding sequence fo
c	460	17	0.3	877	11	Q03999 Sequence complemen
	461	17	0.3	877	19	V58901 DNA encoding Staph
	462	17	0.3	882	21	Z29253 Rifin 3196 gene.
	463	17	0.3	894	16	Q91491 Zeamatin synergist
	464	17	0.3	900	17	T07264 Mutant LAG1 coding
	465	17	0.3	927	20	X02912 B. alkalophilus be
	466	17	0.3	937	19	X14301 H. pylori GHPO 66
	467	17	0.3	941	20	V84618 Human secreted pro
	468	17	0.3	941	20	V84413 Human secreted pro
	469	17	0.3	972	20	X04341 Human secreted pro
	470	17	0.3	972	21	Z65085 Membrane-bound pro
	471	17	0.3	975	19	V61499 Bak binding protei
	472	17	0.3	1010	20	X13442 Enterococcus faeca
	473	17	0.3	1014	21	Z50349 Human microvascula
	474	17	0.3	1016	21	A09027 Human CSAPTP-2 cod
	475	17	0.3	1021	21	A26441 Human secreted pro
	476	17	0.3	1025	21	Z98242 Human signal pepti
	477	17	0.3	1046	12	Q10377 Plasmid pMG3C9 use
	478	17	0.3	1047	20	Z41305 Human normal ovari
c	479	17	0.3	1098	21	Z45565 Nucleotide sequenc
	480	17	0.3	1100	17	T11067 Marrow stromal cel
	481	17	0.3	1120	20	Z33531 Human prostate can
	482	17	0.3	1122	21	Z53961 Neisseria gonorrhe
	483	17	0.3	1125	7	N60273 Sequence encoding
	484	17	0.3	1131	20	X02219 M. piperita gerany
	485	17	0.3	1131	21	Z52306 Mint geranyl dipho
	486	17	0.3	1138	19	V29785 Galectin 8 cDNA.
c	487	17	0.3	1141	11	Q05690 Serum somatotropin
c	488	17	0.3	1150	20	X13878 Enterococcus faeca
	489	17	0.3	1151	8	N70263 Sequence of nitrat
	490	17	0.3	1151	8	N70660 Sequence of the bN
	491	17	0.3	1172	20	X85052 Human secreted pro
	492	17	0.3	1174	20	X97869 Human secreted pro
	493	17	0.3	1182	20	Z12351 Neisseria meningit
	494	17	0.3	1183	20	Z77553 Human ovarian tumo
	495	17	0.3	1198	20	Z21098 Human secreted pro
c	496	17	0.3	1211	14	Q37716 K1ADH4 promoter.
c	497	17	0.3	1212	20	Z28298 Rat neuronal immed
	498	17	0.3	1216	16	Q87295 Rat regucalcin cDN
c	499	17	0.3	1231	20	Z28265 Rat neuronal immed
	500	17	0.3	1232	19	Z96438 S. pneumoniae deri

c	501	17	0.3	1233	20	X91663 Porphorymonas ging
	502	17	0.3	1255	20	X37455 Human secreted pro
	503	17	0.3	1279	21	Z51790 Wheat farnesyltran
c	504	17	0.3	1284	20	X91542 Porphorymonas ging
	505	17	0.3	1290	20	Z00447 Human secreted pro
	506	17	0.3	1308	20	X22114 Human secreted pro
	507	17	0.3	1326	15	Q65606 Porcine zona pellu
	508	17	0.3	1326	20	Z27711 Porcine zona pellu
	509	17	0.3	1326	20	V64788 Porcine ZPC cDNA.
	510	17	0.3	1326	21	Z95648 Porcine zona pellu
	511	17	0.3	1326	21	Z46256 Porcine oocyte zon
	512	17	0.3	1326	21	Z33245 Porcine zona pellu
	513	17	0.3	1326	21	Z37800 Porcine zona pellu
c	514	17	0.3	1369	15	Q66073 Sequence of castor
c	515	17	0.3	1382	14	Q37053 pGPD-1(HBS) fragme
	516	17	0.3	1395	21	A16665 Human secreted pro
c	517	17	0.3	1400	20	Z20368 Gene encoding bact
	518	17	0.3	1403	20	X86970 Truncated plastid
	519	17	0.3	1413	19	V26024 Human ciao-1 gene.
	520	17	0.3	1430	16	Q92525 Nicotiana plumbagi
	521	17	0.3	1433	20	X97822 Human secreted pro
	522	17	0.3	1454	19	V23913 Plant CAD enzyme D
	523	17	0.3	1454	20	Z06879 Pine cinnamyl alco
	524	17	0.3	1468	20	X13315 Enterococcus faeca
c	525	17	0.3	1471	20	Z33548 Human breast tumou
	526	17	0.3	1474	19	V23914 Plant CAD enzyme D
	527	17	0.3	1474	20	Z06880 Pine cinnamyl alco
	528	17	0.3	1475	20	X90838 Maize histone deac
	529	17	0.3	1475	20	X97650 Extended human sec
	530	17	0.3	1511	17	T29032 Endoglucanase (58
	531	17	0.3	1537	19	V40520 Homo sapiens BP202
	532	17	0.3	1540	21	Z65253 Human secreted pro
	533	17	0.3	1542	20	Z11910 Human potassium ch
c	534	17	0.3	1543	20	Z08791 Human Ras protein
	535	17	0.3	1543	20	Z00466 Human secreted pro
	536	17	0.3	1548	21	Z98142 Human signal pepti
	537	17	0.3	1562	14	Q42769 Ligand-induced gen
	538	17	0.3	1576	13	Q25339 Chicken glucagon c
	539	17	0.3	1582	21	Z52548 Human secreted pro
c	540	17	0.3	1587	20	X13606 Enterococcus faeca
c	541	17	0.3	1589	21	Z95198 Human UGT2B4 exon
	542	17	0.3	1593	19	V23274 Human gp49 HM43 en
c	543	17	0.3	1604	21	Z51983 Mouse Phox2b nucle
	544	17	0.3	1620	19	V38992 Human monocyte gen
	545	17	0.3	1630	18	T90432 Senna alpha-galact
	546	17	0.3	1636	14	Q38415 Polygalacturonase
	547	17	0.3	1637	9	N80487 Plasmid clone pTom
	548	17	0.3	1638	20	X52261 Protein PRO307 cDN
c	549	17	0.3	1641	18	T92600 Esterase gene EST2
	550	17	0.3	1641	20	X55569 Seq ID No: 5 of WO
	551	17	0.3	1642	19	V21213 Human type II GnRH

	552	17	0.3	1643	20	X08522 NBP46 (root lectin
	553	17	0.3	1677	21	Z53962 Neisseria meningit
	554	17	0.3	1677	21	Z53963 Neisseria meningit
	555	17	0.3	1680	9	N80488 Plasmid clone pPE1
	556	17	0.3	1680	14	Q38416 Pectin esterase cD
	557	17	0.3	1684	19	V34398 Borage delta-6 des
	558	17	0.3	1684	20	X24917 Borage delta-6 des
	559	17	0.3	1685	17	T30395 Borage delta-6-des
	560	17	0.3	1690	16	Q92524 Nicotiana alata ar
	561	17	0.3	1690	20	X04614 Genomic sequence e
	562	17	0.3	1703	19	V09774 Human marcoSR cDNA
	563	17	0.3	1705	20	X97966 Human secreted pro
	564	17	0.3	1726	20	X07570 Homo sapiens fetal
	565	17	0.3	1750	20	X58404 Tobacco fatty acid
	566	17	0.3	1751	20	Z11909 Human potassium ch
c	567	17	0.3	1758	19	V58786 Human phospholipas
	568	17	0.3	1764	20	X86117 cDNA encoding a Tr
	569	17	0.3	1766	21	A26299 Human secreted pro
c	570	17	0.3	1767	20	X99572 Nucleic acid seque
	571	17	0.3	1783	12	Q15136 Clone ptACC3 encod
	572	17	0.3	1786	18	T71716 Human integrin-lin
	573	17	0.3	1789	18	T94655 Petunia flavonoid
	574	17	0.3	1800	20	Z11907 Human potassium ch
	575	17	0.3	1801	17	T29602 Human EHOC-1 3' un
	576	17	0.3	1806	17	T31094 Candida albicans L
	577	17	0.3	1809	18	T79353 Staphylococcus aur
	578	17	0.3	1809	20	X32452 Plasmid pSJ8-2a pa
	579	17	0.3	1816	21	Z50703 p70(beta) S6 kinas
	580	17	0.3	1821	20	V84443 Human secreted pro
	581	17	0.3	1836	20	Z11908 Human potassium ch
	582	17	0.3	1841	20	Z27262 Human secreted pro
c	583	17	0.3	1864	18	T62961 Human c-Fos induce
	584	17	0.3	1866	20	Z42004 Human endometrium
	585	17	0.3	1875	17	T41614 Adzuki bean endo-x
	586	17	0.3	1883	18	V06639 Arabidopsis thalia
	587	17	0.3	1929	21	Z93789 Human MDG1 gene (C
	588	17	0.3	1947	20	X61429 DNA encoding a hum
	589	17	0.3	1954	21	A23441 cDNA encoding huma
	590	17	0.3	1965	17	T41615 Adzuki bean endo-x
	591	17	0.3	1965	17	T29034 Endoglucanase (nov
c	592	17	0.3	1983	20	V69726 Tumour rejection a
	593	17	0.3	1985	20	V84497 Human secreted pro
	594	17	0.3	1999	18	T93250 Beta-2 adrenalin r
	595	17	0.3	2004	10	N91031 Glycoprotein A ant
	596	17	0.3	2059	21	Z46147 cDNA sequence enco
	597	17	0.3	2104	13	Q25273 Sequence encoding
	598	17	0.3	2134	17	T07263 LAG1 coding sequen
c	599	17	0.3	2190	20	Z24407 Human bladder tumo
	600	17	0.3	2198	20	X86973 Plastid form of a
	601	17	0.3	2230	19	V15704 Tomato ACC synthas
	602	17	0.3	2235	20	Z11208 Mutant threonine d

	603	17	0.3	2241	20	Z11206 Mutant threonine d
	604	17	0.3	2248	19	V07918 Helicobacter pylor
c	605	17	0.3	2260	17	T27730 Insulin-stimulated
c	606	17	0.3	2260	17	T27731 Mutant insulin-sti
	607	17	0.3	2262	21	Z52560 Human secreted pro
	608	17	0.3	2271	21	Z99964 DNA encoding a hum
	609	17	0.3	2274	19	V70285 Human osteoprotege
	610	17	0.3	2277	20	Z11212 Mutant threonine d
	611	17	0.3	2280	21	Z36947 cDNA encoding a MP
	612	17	0.3	2306	16	Q91290 ORC5 gene encoding
	613	17	0.3	2306	18	T73283 S. cerevisiae orig
	614	17	0.3	2338	19	V22743 Babesia microti BM
	615	17	0.3	2338	20	X90030 Rat connective tis
	616	17	0.3	2338	20	X90007 Babesia microti an
	617	17	0.3	2360	15	Q54840 Sequence encoding
	618	17	0.3	2383	17	T11418 p53 response prote
	619	17	0.3	2397	9	N81441 Clone encoding bov
	620	17	0.3	2406	18	T43376 Human cytokine res
	621	17	0.3	2443	20	V73477 Human rab7GTP anal
	622	17	0.3	2481	19	V40019 Mouse Pax6 gene
	623	17	0.3	2481	19	V40035 Mouse Pax6 gene SE
	624	17	0.3	2481	20	Z10930 Pax6 coding sequen
c	625	17	0.3	2511	17	T34685 Squash mosaic viru
c	626	17	0.3	2547	11	Q05691 Serum somatotropin
c	627	17	0.3	2574	19	V63209 Nucleic acid seque
	628	17	0.3	2616	20	Z12353 Neisseria meningit
	629	17	0.3	2667	20	Z12352 Neisseria meningit
	630	17	0.3	2690	20	Z34317 Human PRO298 nucle
	631	17	0.3	2694	21	Z96806 Nuclear transport
	632	17	0.3	2729	17	T41469 Mouse 030 gene (fo
c	633	17	0.3	2755	21	A09173 Human zrnpl (ribon
	634	17	0.3	2759	21	Z33372 Human secreted pro
	635	17	0.3	2762	20	Z00049 Human growth facto
	636	17	0.3	2766	21	Z61582 DNA encoding a BAS
	637	17	0.3	2769	20	Z12354 Neisseria gonorrho
	638	17	0.3	2769	21	Z61580 DNA encoding a BAS
	639	17	0.3	2769	21	Z61581 DNA encoding a BAS
	640	17	0.3	2784	12	Q11701 Human alpha-interf
c	641	17	0.3	2785	15	Q69000 Product of alterna
c	642	17	0.3	2785	17	T03624 Alternatively spli
	643	17	0.3	2811	20	X05817 Rat pheromone rece
	644	17	0.3	2839	11	Q03742 Human SKI related
	645	17	0.3	2940	14	Q33473 HD1(T7). Human im
	646	17	0.3	2940	20	V08705 HD1(T7) DNA constr
c	647	17	0.3	2940	20	V69727 Tumour rejection a
	648	17	0.3	2981	19	V70360 Human ehb21 encodi
	649	17	0.3	3032	19	V32790 DNA sequence encod
	650	17	0.3	3043	18	T85407 Rat truncated cadh
	651	17	0.3	3043	18	T61918 Truncated rat cadh
	652	17	0.3	3091	20	X51707 DNA encoding a hum
	653	17	0.3	3138	12	Q11712 Shuttle vector pMU

c	654	17	0.3	3147	21	X82996 P. falciparum gcpE
	655	17	0.3	3153	16	Q88241 Neisseria gonorrho
	656	17	0.3	3178	20	X99865 Human secreted pro
	657	17	0.3	3192	11	Q06055 5'sequence of yeas
	658	17	0.3	3195	20	X36712 Human phosphodiest
	659	17	0.3	3312	15	Q79326 Human ARD 1. Homo
c	660	17	0.3	3335	10	N91154 Turkey Herpes Viru
c	661	17	0.3	3335	17	T31898 Herpesvirus of tur
	662	17	0.3	3336	17	T35912 Herpesvirus of tur
	663	17	0.3	3375	21	A35101 Human adenosine re
c	664	17	0.3	3395	14	Q42759 IBDV gene sequence
	665	17	0.3	3494	20	X23729 Human CD13/aminope
c	666	17	0.3	3496	19	V32554 Candida albicans C
	667	17	0.3	3542	19	V59563 Human secreted pro
c	668	17	0.3	3764	21	Z44192 Murine cerebral ne
	669	17	0.3	3785	18	T79275 Human beta 2 integ
	670	17	0.3	3785	19	V67299 Human alpha-d deri
	671	17	0.3	3785	19	V35316 Human Beta-integri
	672	17	0.3	3785	20	V08458 Human alphad codin
	673	17	0.3	3934	19	T97233 Trimmed enzyme enc
	674	17	0.3	3956	18	T79274 Human beta 2 integ
	675	17	0.3	3956	19	V67298 Human alpha-d clon
	676	17	0.3	3956	19	V35315 Human Beta-integri
	677	17	0.3	3956	20	V08457 Human alphad codin
c	678	17	0.3	3958	17	T43596 Glutamate-gated ch
	679	17	0.3	4166	21	Z36816 DNA encoding a sev
c	680	17	0.3	4170	21	Z99468 Arabidopsis gibber
	681	17	0.3	4196	20	Z40488 Human SULU1 DNA.
c	682	17	0.3	4233	20	X87946 Candida tropicalis
	683	17	0.3	4322	19	V58192 Mouse Bsk receptor
	684	17	0.3	4434	20	X32300 Fumonosin-resistan
c	685	17	0.3	4467	18	T68648 pSRQ800 fragment i
c	686	17	0.3	4471	20	X00915 Protease activated
	687	17	0.3	4472	16	T05406 Human cell cycle g
	688	17	0.3	4534	20	X03042 Human IL-1ra BAC c
c	689	17	0.3	4590	7	N60472 Sequence encoding
c	690	17	0.3	4599	17	T27052 RRP3 telomerase-as
c	691	17	0.3	4705	15	Q68998 Human protocadheri
c	692	17	0.3	4705	17	T03622 Protocadherin clon
c	693	17	0.3	4976	11	Q05754 Microspore-specifi
c	694	17	0.3	4977	18	T48863 Brassica napus mic
	695	17	0.3	5020	19	V52275 Streptococcus pneu
	696	17	0.3	5108	7	N60388 Sequence encoding
	697	17	0.3	5150	18	T78598 AS-30D tumour Type
	698	17	0.3	5170	21	A16683 Human secreted pro
c	699	17	0.3	5263	21	A16627 Human secreted pro
	700	17	0.3	5340	16	Q75956 S. cerevisiae scau
c	701	17	0.3	5347	20	Z09633 B. lactoferamentum
	702	17	0.3	5426	17	T18210 Turkey Herpes Viru
	703	17	0.3	5457	20	X35733 cDNA encoding a pr
	704	17	0.3	5468	11	Q06540 Human low density

	705	17	0.3	5532	20	Z25100 Human ischaemic he
c	706	17	0.3	5760	6	N50530 Sequence encoding
	707	17	0.3	6046	19	V13845 Complete DNA seque
c	708	17	0.3	6152	18	T78867 P. falciparum live
	709	17	0.3	6157	18	V74495 Staphylococcus aur
c	710	17	0.3	6202	16	Q88161 Q8 ethylene respon
c	711	17	0.3	6202	19	V59040 A. thaliana ethyle
	712	17	0.3	6245	19	V13844 Complete DNA seque
	713	17	0.3	6247	20	X86997 A. thaliana hydroc
c	714	17	0.3	6346	18	V74548 Staphylococcus aur
	715	17	0.3	6448	19	V13843 Complete DNA seque
	716	17	0.3	6580	19	V13167 Complete DNA seque
	717	17	0.3	6613	19	V13842 Complete DNA seque
	718	17	0.3	6710	20	X90998 Human centrin-2 ge
c	719	17	0.3	6763	18	T79728 Beta-lactoglobulin
	720	17	0.3	6771	21	A35103 Human adenosine re
	721	17	0.3	6796	13	Q27433 HVTA antigen (gA)
	722	17	0.3	6869	18	T47442 Saccharomyces cere
	723	17	0.3	6959	19	V13165 Complete DNA seque
	724	17	0.3	6995	19	V13162 Complete DNA seque
	725	17	0.3	7002	19	V13166 Complete DNA seque
	726	17	0.3	7102	20	X31836 Nucleotide sequenc
	727	17	0.3	7207	21	A35194 Human adenosine re
c	728	17	0.3	7227	13	Q25183 E.faecium antibiot
	729	17	0.3	7333	20	X31837 Nucleotide sequenc
c	730	17	0.3	7647	19	V15519 Hamster oral papil
	731	17	0.3	7859	12	Q12154 Shuttle vector pSW
c	732	17	0.3	7859	12	Q12154 Shuttle vector pSW
	733	17	0.3	7859	14	Q44265 pSW6 for expressio
c	734	17	0.3	7859	14	Q44265 pSW6 for expressio
	735	17	0.3	7984	13	Q25185 pSW6 expression ve
c	736	17	0.3	7984	13	Q25185 pSW6 expression ve
	737	17	0.3	8365	20	Z20056 Plasmodium falcipa
	738	17	0.3	8501	19	V69925 Leukotoxin activat
	739	17	0.3	8536	18	V74589 Staphylococcus aur
c	740	17	0.3	8639	20	X02995 Human IL-1ra BAC c
	741	17	0.3	8760	15	Q73473 Porcine pro-interl
c	742	17	0.3	9060	12	Q15132 Zucchini ACC synth
c	743	17	0.3	9060	19	V15702 Zucchini ACC synth
c	744	17	0.3	9180	19	V57416 Tumour suppressor
	745	17	0.3	9603	20	X60894 DHFR-TS gene seque
	746	17	0.3	9705	11	Q02828 Complete genomic s
c	747	17	0.3	10288	18	T71322 Plasmid pCB51 enco
c	748	17	0.3	10807	16	T03855 Sheep beta-lactogl
c	749	17	0.3	10807	18	T79725 Ovine beta-lactogl
c	750	17	0.3	10851	13	Q25178 E.faecium antibiot
	751	17	0.3	10968	19	V17789 Tomato Prf genomic
	752	17	0.3	11958	19	V22961 Plasmid BV-C2PG en
c	753	17	0.3	12124	20	V62854 Cucumber chitinase
c	754	17	0.3	12124	20	V81625 Cucumber chitinase
	755	17	0.3	13330	20	Z23893 Human LOBO genomic

c 756	17	0.3	13414	18	T71321	Plasmid pCB50 enco
757	17	0.3	13500	21	Z45562	Nucleotide sequenc
c 758	17	0.3	13574	20	X13051	Enterococcus faeca
759	17	0.3	14872	19	V52205	Streptococcus pneu
760	17	0.3	15630	21	A35195	Human adenosine re
c 761	17	0.3	16826	18	V74357	Staphylococcus aur
762	17	0.3	16956	21	Z91923	Wild type (C57BL/6
763	17	0.3	17041	13	Q21065	Genomic DNA of hum
764	17	0.3	17041	15	Q67057	PACPAP38 DNA. Hom
c 765	17	0.3	19403	20	X20541	Polynucleotide seq
c 766	17	0.3	20860	20	X13292	Enterococcus faeca
c 767	17	0.3	22191	20	X20600	Polynucleotide seq
768	17	0.3	30600	20	Z32023	Human METH1 relate
769	17	0.3	35414	21	D00147	TR12 related DNA-1
770	17	0.3	37856	21	A11992	S. cellulosum DNA
c 771	17	0.3	38186	20	Z32028	Human METH1 relate
772	17	0.3	38644	21	A35302	Human adenosine re
773	17	0.3	43676	21	Z60606	Nucleotide sequenc
774	17	0.3	45546	20	X23520	Human kidney amino
775	17	0.3	49999	20	Z23900	Human LOBO homolog
c 776	17	0.3	68750	21	Z55887	Sorangium cellulos
777	17	0.3	72604	20	Z10752	Genomic sequence o
c 778	17	0.3	72604	20	Z10752	Genomic sequence o
779	17	0.3	80240	20	V83940	NC-contig derived
780	17	0.3	80595	20	V83939	NC-contig derived
c 781	17	0.3	87350	18	X83003	Human WRN genomic
c 782	17	0.3	90050	21	Z91925	Wild type (C57BL/6
783	17	0.3	143068	21	A34983	Human adenosine re
784	17	0.3	143068	21	A35150	Human adenosine re
c 785	17	0.3	144460	21	Z93815	Olfactory receptor
786	17	0.3	149412	21	A35151	Human adenosine re
787	17	0.3	580073	18	T58840	Mycoplasma genital
c 788	17	0.3	1664976	19	V21209	Methanococcus jann
789	17	0.3	1830121	17	T42063	Haemophilus influe
c 790	16	0.3	16	20	X18362	RT-PCR primer of t
c 791	16	0.3	18	19	V54166	Nucleotide sequenc
c 792	16	0.3	18	21	Z90648	Human adipose tiss
c 793	16	0.3	19	16	Q75556	Reverse transcript
c 794	16	0.3	20	14	Q41648	pKPIAV expression
c 795	16	0.3	20	16	Q75596	Reverse transcript
c 796	16	0.3	20	16	Q75598	Reverse transcript
c 797	16	0.3	20	16	Q75595	Reverse transcript
c 798	16	0.3	21	16	Q75751	Reverse transcript
c 799	16	0.3	21	16	Q75752	Reverse transcript
c 800	16	0.3	21	16	Q75753	Reverse transcript
c 801	16	0.3	21	16	Q75754	Reverse transcript
c 802	16	0.3	21	16	Q75755	Reverse transcript
c 803	16	0.3	21	16	Q75756	Reverse transcript
c 804	16	0.3	21	16	Q75757	Reverse transcript
c 805	16	0.3	21	16	Q75758	Reverse transcript
c 806	16	0.3	21	16	Q75763	Reverse transcript

c 807	16	0.3	21	16	Q75764	Reverse transcript
c 808	16	0.3	21	16	Q75765	Reverse transcript
c 809	16	0.3	21	16	Q75766	Reverse transcript
810	16	0.3	21	19	Z26235	Human polymorphic
811	16	0.3	26	21	Z49474	Human Butyrylcholi
c 812	16	0.3	27	20	X23568	Deletion sequence
813	16	0.3	28	21	Z55421	Neisseria species
c 814	16	0.3	30	18	T69677	Downstream primer
815	16	0.3	30	20	X34263	Forward primer for
816	16	0.3	31	19	V48095	Oligonucleotide 31
817	16	0.3	34	19	V48094	Oligonucleotide 34
c 818	16	0.3	35	14	Q48774	Aphrodisine primer
c 819	16	0.3	35	17	T17245	Human beta-defensi
c 820	16	0.3	36	18	T48492	Bax omega protein
c 821	16	0.3	36	20	X87557	Pichia pastoris al
822	16	0.3	37	19	V48093	Oligonucleotide 37
c 823	16	0.3	38	21	Z43896	M. tuberculosis rp
c 824	16	0.3	38	21	Z43897	M. tuberculosis rp
c 825	16	0.3	40	13	Q25020	Sense oligonucleot
c 826	16	0.3	40	13	Q25023	Anti-sense oligonu
c 827	16	0.3	40	18	T61448	Nucleotide anchore
c 828	16	0.3	40	19	V47827	Maize polymorphic
829	16	0.3	40	19	V48092	Oligonucleotide 40
c 830	16	0.3	40	21	Z59458	PCR primer RP10.3
831	16	0.3	40	21	Z95793	Polynucleotide seq
c 832	16	0.3	42	19	V27042	Primer PLHNAP-REV.
833	16	0.3	43	19	V48091	Oligonucleotide 43
834	16	0.3	50	11	Q06082	Capture probe for
c 835	16	0.3	51	21	Z92266	Substrate oligonuc
836	16	0.3	51	21	Z44863	Construct pPAMut2
837	16	0.3	54	18	T73959	Knot probe P302 us
838	16	0.3	55	18	V76122	Staphylococcus aur
839	16	0.3	55	18	T72962	Omega probe P302 u
c 840	16	0.3	63	20	X58042	Primer for rat GAB
c 841	16	0.3	63	21	Z51428	PCR primer MJ-47 t
c 842	16	0.3	65	21	Z42904	Human 5' EST isola
c 843	16	0.3	69	18	T84542	KIRK-2 gene antise
844	16	0.3	80	16	Q75103	Plasmid pOKSC8a co
845	16	0.3	97	17	T39460	Growth regulatory
846	16	0.3	97	19	V37478	Human growth regul
847	16	0.3	109	17	T39454	Growth regulatory
848	16	0.3	109	19	V37472	Human growth regul
c 849	16	0.3	115	16	T20719	Human gene signatu
850	16	0.3	115	17	T39461	Growth regulatory
851	16	0.3	115	19	V37479	Human growth regul
852	16	0.3	144	16	T25320	Human gene signatu
853	16	0.3	157	20	Z14407	Human gene express
c 854	16	0.3	165	18	V75693	Staphylococcus aur
855	16	0.3	175	20	Z14983	Human gene express
856	16	0.3	177	18	T91297	Human G52-24 secre
857	16	0.3	177	18	V12823	3' fragment of clo

858	16	0.3	180	16	T35104	Down-regulated sen
859	16	0.3	180	18	T67738	H. pylori cytoplas
860	16	0.3	180	18	T77418	H. pylori cytoplas
861	16	0.3	190	21	Z47279	C-fos/biotinylatio
862	16	0.3	190	21	Z56670	C-fos fusion partn
863	16	0.3	197	18	T89603	Hepatitis C virus
864	16	0.3	199	14	Q42772	Ligand-induced gen
865	16	0.3	199	19	X12112	Human biallelic po
c 866	16	0.3	203	15	Q76439	Human genome fragm
867	16	0.3	212	20	Z14371	Human gene express
868	16	0.3	233	20	Z14370	Human gene express
869	16	0.3	245	19	V30925	Human secreted pro
870	16	0.3	246	17	T11863	L. ivanovii CIP 78
871	16	0.3	246	19	V37716	Mouse protein L256
872	16	0.3	246	20	V88799	EST clone H0693.
873	16	0.3	247	14	Q59958	Human brain Expres
874	16	0.3	254	18	X83276	Breast cancer tumo
875	16	0.3	254	19	V68874	DNA molecule encod
876	16	0.3	257	20	Z24548	Human lung tumor a
c 877	16	0.3	258	14	Q61367	Human brain Expres
878	16	0.3	258	17	T37134	Proinsulin coding
879	16	0.3	262	18	T75501	P. americanus skin
880	16	0.3	270	18	T67889	H. pylori cytoplas
881	16	0.3	270	21	A00369	Human colon cancer
c 882	16	0.3	271	11	Q03702	Fragment of plasmi
c 883	16	0.3	272	11	Q02477	Phosphoglycerate k
884	16	0.3	275	13	Q29235	Sequence encoding
885	16	0.3	275	16	T04834	Diguetia mature to
886	16	0.3	275	18	T85940	Deguitia spider ve
887	16	0.3	275	18	T89372	DNA encoding Digue
888	16	0.3	279	20	Z24543	Human lung tumor a
c 889	16	0.3	280	20	V89654	EST clone CU32. H
890	16	0.3	284	20	Z24580	Human lung tumor a
891	16	0.3	285	18	T92843	Flea saliva protei
892	16	0.3	285	19	V73391	Flea saliva protei
c 893	16	0.3	286	16	T20372	Human gene signatu
c 894	16	0.3	289	20	Z13670	Human gene express
c 895	16	0.3	296	21	A00383	Human colon cancer
896	16	0.3	300	20	Z14372	Human gene express
897	16	0.3	300	20	Z13988	Human gene express
898	16	0.3	300	20	Z12930	Human gene express
899	16	0.3	300	20	X98485	Human cancer cell
900	16	0.3	300	20	X98463	Human cancer cell
901	16	0.3	301	20	X40522	Human secreted pro
c 902	16	0.3	301	21	A05514	Human immunogenic
903	16	0.3	303	18	T68120	H. pylori transmem
904	16	0.3	303	20	V90279	EST clone D1387.
905	16	0.3	305	13	Q32653	LEMMI 2 cDNA. Lyc
906	16	0.3	306	18	T94580	Cloned 3' region o
907	16	0.3	315	20	Z24501	Human lung tumor a
908	16	0.3	324	20	X34125	Mycobacterium spec

909	16	0.3	325	20	X34124	Mycobacterium spec
c 910	16	0.3	325	20	V90511	EST clone DW705.
c 911	16	0.3	326	16	T22639	Human gene signatu
912	16	0.3	326	20	X34126	Mycobacterium spec
913	16	0.3	328	14	Q60582	Human brain Expres
914	16	0.3	330	16	T05576	Encodes human citr
915	16	0.3	336	18	T68122	H. pylori outer me
916	16	0.3	336	20	V86586	EST clone BA210.
917	16	0.3	339	11	Q06879	Encodes human biku
918	16	0.3	340	16	T19982	Human gene signatu
919	16	0.3	341	18	T63527	Pichia pastoris 3'
920	16	0.3	341	19	X14470	H. pylori GHPO 104
c 921	16	0.3	342	14	Q59382	Human brain Expres
c 922	16	0.3	344	14	Q60243	Human brain Expres
923	16	0.3	345	18	T67750	H. pylori outer me
924	16	0.3	345	18	T77430	H. pylori outer me
c 925	16	0.3	345	20	V86768	EST clone AY307.
926	16	0.3	349	14	Q59398	Human brain Expres
927	16	0.3	349	20	X40387	Human secreted pro
c 928	16	0.3	352	20	Z09014	Differentiation In
929	16	0.3	354	21	Z93803	Sequence encoding
930	16	0.3	360	18	T88078	3' portion of cDNA
931	16	0.3	360	18	V02298	Human secreted pro
932	16	0.3	360	18	V02140	Human secreted pro
933	16	0.3	360	20	X27421	Human secreted pro
934	16	0.3	360	21	Z29601	Rat cDNA clone rG5
935	16	0.3	361	14	Q36483	Mycoplasma primer/
936	16	0.3	361	20	X41278	Human secreted pro
c 937	16	0.3	362	16	T23217	Human gene signatu
938	16	0.3	363	17	T11865	L. ivanovii CIP 78
939	16	0.3	367	20	X84366	Drosophila Acp33A
940	16	0.3	367	21	Z80362	Human colon cancer
941	16	0.3	374	19	X14290	H. pylori GHPO 10
942	16	0.3	375	14	Q59656	Human brain Expres
c 943	16	0.3	379	20	V87480	EST clone BR853.
944	16	0.3	382	19	V38432	cDNA encoding the
c 945	16	0.3	383	19	Z96491	S. pneumoniae deri
c 946	16	0.3	383	20	X77571	S. pneumoniae ribH
947	16	0.3	388	20	X34123	Mycobacterium spec
948	16	0.3	389	20	X34122	Mycobacterium spec
c 949	16	0.3	389	20	V90095	EST clone CY9. Ho
950	16	0.3	390	20	X34121	Mycobacterium spec
951	16	0.3	390	21	A05409	Streptococcus pneu
c 952	16	0.3	392	20	V90111	EST clone DG330..
953	16	0.3	393	14	Q36475	Mycoplasma primer/
954	16	0.3	395	20	Z08306	Human lung tumour
c 955	16	0.3	397	19	X14010	H. pylori GHPO 162
c 956	16	0.3	397	21	A31703	Plant microsatelli
957	16	0.3	398	20	V90231	EST clone DF910.
958	16	0.3	400	18	V78220	Staphylococcus aur
959	16	0.3	400	18	V77401	Staphylococcus aur

c 960	16	0.3	400	18	V76287	Staphylococcus aur
961	16	0.3	400	20	X57380	Rat U3 gene trap d
c 962	16	0.3	409	18	V75559	Staphylococcus aur
c 963	16	0.3	410	14	Q59990	Human brain Expres
964	16	0.3	413	20	X30334	DNA encoding a hum
965	16	0.3	417	14	Q60665	Human brain Expres
c 966	16	0.3	417	20	V89891	EST clone CT358.
967	16	0.3	419	19	V33195	Secreted protein C
968	16	0.3	422	20	V87678	EST clone DO19. H
969	16	0.3	427	20	X61370	DNA encoding a hum
970	16	0.3	428	19	V34202	Human secreted pro
971	16	0.3	428	19	V38430	cDNA encoding the
972	16	0.3	428	19	V16500	DNA encoding a Bac
973	16	0.3	428	19	V16506	DNA encoding a Bac
974	16	0.3	428	19	V16492	DNA encoding a Bac
975	16	0.3	428	20	X83933	Bacillus thuringie
976	16	0.3	428	20	X83927	Bacillus thuringie
977	16	0.3	428	20	X83919	Bacillus thuringie
c 978	16	0.3	431	20	X34034	Mycobacterium spec
c 979	16	0.3	432	20	X34033	Mycobacterium spec
980	16	0.3	433	20	Z24558	Human lung tumor a
c 981	16	0.3	433	20	X34032	Mycobacterium spec
982	16	0.3	436	19	V26703	Human novel secret
983	16	0.3	436	21	A31427	Plant microsatelli
984	16	0.3	437	20	Z21346	Human CBFBMDO4 nuc
985	16	0.3	441	19	V40541	Homo sapiens secre
986	16	0.3	441	20	X20750	Polynucleotide seq
c 987	16	0.3	443	19	V59560	Human secreted pro
c 988	16	0.3	446	14	Q39702	Expressed Sequence
c 989	16	0.3	446	14	Q59114	Human brain Expres
990	16	0.3	448	17	T30571	Eukaryotic cell gr
991	16	0.3	451	20	Z41347	Human normal uteru
992	16	0.3	452	13	Q28007	Sequence encoding
993	16	0.3	452	13	Q28760	Sequence of clone
994	16	0.3	452	20	X97795	Extended human sec
995	16	0.3	452	20	X61375	DNA encoding a hum
996	16	0.3	459	20	X61330	DNA encoding a hum
c 997	16	0.3	462	20	V89672	EST clone DA274.
998	16	0.3	464	19	V38682	Mus musculus SOCS1
999	16	0.3	465	20	X37518	Human secreted pro
1000	16	0.3	468	20	X77570	S. pneumoniae ribH

ALIGNMENTS

```
RESULT   1
Q48733
ID   Q48733 standard; DNA; 5925 BP.
XX
AC   Q48733;
```



```
XX
DT   17-MAR-1994  (first entry)
XX
DE   CAI gene.
XX
KW   Cytotoxin; CT; H. pylori; precursor; vacuolation; cell death;
KW   eukaryote; cytotoxin-associated immunodominant antigen; CAI;
KW   heat shock protein; hsp; hsp60; type B; gastritis; peptic ulcer;
KW   gastric tumours; ss.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             535..3978
FT                   /*tag= a
FT                   /product= CAI antigen
FT   RBS             525..529
FT                   /*tag= b
FT   terminator      4010..4032
FT                   /*tag= c
FT                   /standard_name= Rho_independant_terminator
XX
PN   WO9318150-A.
XX
PD   16-SEP-1993.
XX
PF   02-MAR-1993;   93WO-EP00472.
XX
PR   02-MAR-1992;   92IT-0FI0052.
PR   25-JAN-1993;   93WO-EP00158.
XX
PA   (BIOC-) BIOCINE SCLAVO SPA.
XX
PI   Bugnoli M,  Covacci A,  Macchia G,  Rappuoli R,  Telford J;
XX
DR   WPI; 1993-303464/38.
DR   P-PSDB; R41199.
XX
PT   Recombinant Helicobacter pylori protein and corresp. gene - is a
PT   cytotoxin, antigen or heat shock protein used for treating and
PT   preventing type B gastritis, gastric ulcers and gastric tumours
XX
PS   Claim 28; Fig 4; 83pp; English.
XX
CC   This sequence encodes the cytotoxin-associated immunodominant antigen,
CC   CAI, of H. pylori.  The protein is a hydrophillic, surface-exposed
CC   protien having a molecular weight of approximately 120-132 kD, pref.
CC   128-130 kD, and an isoleletric point of 9.72.  H. pylori isolates
CC   which do not produde the CAI antigen do not have the cai gene, and
CC   are also unable to produce an active cytotoxin.  The association
CC   between the presence of the cai gene and cytotoxicity suggests that
```



```
CC   the product of the cai gene is necessary for the transcription,
CC   folding, export or function of the cytotoxin.  Alternatively both
CC   the cai gene and the ct gene are absent in noncytotoxic strains
CC   suggesting physical linkage between the genes.  The absence of a
CC   typical leader sequence suggests the presence of an independant export
CC   system.  Computer searches for promoter regions in the region upstream
CC   from the ATG codon identified sequences resembling either -10 or -35
CC   regions, however a region with a good consensus to the E. coli or H.
CC   pylori promoter sequences was not found.  The CAI antigen is very
CC   hydrophilic and does not show obvious leader peptide or transmembrane
CC   sequences.  The most hydrophilic region is from amino acids 600-900,
CC   where a number of unusual features are observed.  There is a
CC   repetition of the sequences EFKNGKNKDPSK and EPYIA, and the presence
CC   of a stretch of six contiguous asparagines.  This protein, and others
CC   derived from H. pylori, esp. cytotoxin (CT) or a heat shock protein
CC   (see also Q48732 and Q48734), may be used to treat, prevent and
CC   diagnose H. pylori infection.  H. pylori is the causative agent of
CC   type B gastritis, peptic ulcers and gastric tumours.
XX
SQ   Sequence 5925 BP; 2076 A; 1057 C; 1138 G; 1654 T; 0 other;

  Query Match           100.0%;  Score 5925;  DB 14;  Length 5925;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 5925;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctccatttaagcaactccatagaccactaaagaaactttttttgaggctatctttgaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ctccatttaagcaactccatagaccactaaagaaactttttttgaggctatctttgaaa 60

Qy      61 atctgtcctattgatttgttttccatttttgtttcccatgtggatcttgtggatccacaaac 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 atctgtcctattgatttgttttccatttttgtttcccatgtggatcttgtggatccacaaac 120

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgattttct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgattttct 240

Qy     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttttagcaaattt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttttagcaaattt 300

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360

Qy     361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaaatgcttgatattgttgtat 420
```



```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaaatgcttgatattgttgtat 420

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480

Qy     481 ttgacactaacaagataccgataggtatgaaactaggtatagtaaggagaaaacaatgact 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ttgacactaacaagataccgataggtatgaaactaggtatagtaaggagaaaacaatgact 540

Qy     541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600

Qy     601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660

Qy     661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720

Qy     721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780

Qy     781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840

Qy     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 900

Qy     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatg 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatg 960

Qy     961 gaaaatatcatacaacccccctatccttgatgataaagagaaagcggagtttttgaaatct 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 gaaaatatcatacaacccccctatccttgatgataaagagaaagcggagtttttgaaatct 1020

Qy    1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080

Qy    1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140
```

```
Qy   1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200

Qy   1201 gatgtcaaagaagcaatcaatcaagaacccagttccccatgtccaacccagatatagccaact 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 gatgtcaaagaagcaatcaatcaagaacccagttccccatgtccaacccagatatagccaact 1260

Qy   1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320

Qy   1321 aatttttctaaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 aatttttctaaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380

Qy   1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440

Qy   1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500

Qy   1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740

Qy   1741 cttgcacaaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 cttgcacaaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800

Qy   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860

Qy   1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
```



```
Qy   1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980

Qy   1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040

Qy   2041 tattctaatttcaaatacaccaacgcctccaagaatcccaatgaagggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 tattctaatttcaaatacaccaacgcctccaagaatcccaatgaagggtgtaggcgttacg 2100

Qy   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160

Qy   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220

Qy   2221 aaaggattgtcccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 aaaggattgtcccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280

Qy   2281 gaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 gaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340

Qy   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400

Qy   2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460

Qy   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520

Qy   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580

Qy   2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640

Qy   2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagcccctt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagcccctt 2700

Qy   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaaccctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaaccctt 2760

Qy   2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820

Qy   2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880

Qy   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940

Qy   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggccccaacaagct 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggccccaacaagct 3000

Qy   3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060

Qy   3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120

Qy   3121 aaaaactttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 aaaaactttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180

Qy   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300

Qy   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360

Qy   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420

Qy   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600

Qy   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720

Qy   3721 atagttgcgcataatgtaggaagcgttccttttgtcagagtatgataaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 atagttgcgcataatgtaggaagcgttccttttgtcagagtatgataaaattggcttcaac 3780

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840

Qy   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960

Qy   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020

Qy   4021 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttgggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttgggga 4080

Qy   4081 tgaatgctaccaattcatggtatcatatcccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 tgaatgctaccaattcatggtatcatatcccccatacattcgtatctagcgtaggaagtgt 4140

Qy   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
```

```
Qy   4201 aactctgtaaaatccctattataggagacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 aactctgtaaaatccctattataggagacacagagtgagaaccaaactctccctacgggca 4260

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380

Qy   4381 aagacttaacccttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 aagacttaacccttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500

Qy   4501 aagatcaagttcaaaaaatcatagagctttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 aagatcaagttcaaaaaatcatagagctttagagcaaattgatcgcgctcttaaccaaa 4560

Qy   4561 gaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 gaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaaagacgcttgc 4620

Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680

Qy   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860

Qy   4861 tctttagcgctattttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 tctttagcgctattttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920

Qy   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
```



```
Qy   4981 acccttatagtgtgagctatagccccttttttgggaattgagttattttgactttaaatttt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 acccttatagtgtgagctatagccccttttttgggaattgagttattttgactttaaatttt 5040

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100

Qy   5101 gctcgcatgaaattccacttttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 gctcgcatgaaattccacttttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggtgttgaaagggc 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggtgttgaaagggc 5220

Qy   5221 aaaatgctccatcaaaaatagccctcaatttttttgagcgattaagggaaaatgcgtgcaacc 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 aaaatgctccatcaaaaatagccctcaatttttttgagcgattaagggaaaatgcgtgcaacc 5280

Qy   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340

Qy   5341 aacaattcgcccctctaaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 aacaattcgcccctctaaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400

Qy   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460

Qy   5461 gtcgcttttgtccctagcactaaaatagggggcgtttttatctttacttgtcgcttgatc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 gtcgcttttgtccctagcactaaaatagggggcgtttttatctttacttgtcgcttgatc 5520

Qy   5521 gctaaaatgcttggctcaatcacgcccacaatagggattttggaatgcttttgcatctct 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 gctaaaatgcttggctcaatcacgcccacaatagggattttggaatgcttttgcatctct 5580

Qy   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640

Qy   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700

Qy   5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760

Qy   5761 aggctttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctattttcatgaca 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 aggctttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctattttcatgaca 5820

Qy   5821 ctttttttaatttaatgggattaattagggattttattttttcattcattaagtttaaaaat 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 ctttttttaatttaatgggattaattagggattttattttttcattcattaagtttaaaaat 5880

Qy   5881 tcttcattgtccttagtttgttgtcattttagaatagacaaagctt 5925
          |||||||||||||||||||||||||||||||||||||||||||||
Db   5881 tcttcattgtccttagtttgttgtcattttagaatagacaaagctt 5925
```

```
RESULT   2
T46159
ID   T46159 standard; DNA; 19932 BP.
XX
AC   T46159;
XX
DT   30-JUN-1997  (first entry)
XX
DE   CagI locus.
XX
KW   CagI; CagA; virulence factor; exporter molecule; homology; ptl gene;
KW   Bordetella pertussis; VIR B4; Agrobacterium tumefaciens; invasion factor;
KW   Salmonella; type I strain; virulence; diagnosis; H. pylori; infection;
KW   vaccine; treatment; duodenal; gastric ulcer; active gastritis;
KW   adenocarcinoma; as.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             389..1579
FT                   /*tag=  a
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             1794..2168
FT                   /*tag=  b
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             3343..4491
FT                   /*tag=  c
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             4488..5426
FT                   /*tag=  d
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (1778..2173)
```



```
FT                   /*tag=  e
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (2591..3001)
FT                   /*tag=  f
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (3020..3259)
FT                   /*tag=  g
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (5508..6233)
FT                   /*tag=  h
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (6218..6613)
FT                   /*tag=  h
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (6562..7074)
FT                   /*tag=  i
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (7007..7375)
FT                   /*tag=  j
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (7371..7802)
FT                   /*tag=  k
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (7975..8454)
FT                   /*tag=  l
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (8496..8918)
FT                   /*tag=  m
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (8977..9762)
FT                   /*tag=  n
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             complement (9784..10575)
FT                   /*tag=  o
FT                   /note= "putative open reading frame; no start codon
FT                   given"
FT   CDS             10557..12758
FT                   /*tag=  p
FT                   /note= "putative open reading frame; no start codon
FT                   given"
```

```
FT   CDS             complement (12743..13420)
FT                   /*tag= q
FT                   /note= "putative open reading frame; no start codon
FT                        given"
FT   CDS             complement (13374..13742)
FT                   /*tag= r
FT                   /note= "putative open reading frame; no start codon
FT                        given"
FT   CDS             complement (13866..14129)
FT                   /*tag= s
FT                   /note= "putative open reading frame; no start codon
FT                        given"
FT   CDS             complement (19296..19832)
FT                   /*tag= t
FT                   /note= "putative open reading frame; no start codon
FT                        given"
XX
PN   WO9633274-A1.
XX
PD   24-OCT-1996.
XX
PF   18-APR-1996;   96WO-IB00343.
XX
PR   07-JUN-1995;   95US-0477451.
PR   20-APR-1995;   95US-0425194.
XX
PA   (BIOC-) BIOCINE SPA.
XX
PI   Covacci A;
XX
DR   WPI; 1996-485780/48.
DR   P-PSDB; W06930-50.
XX
PT   Helicobacter pylori CagI polynucleotide and related proteins - used
PT   in diagnosis and in vaccines for the treatment of H. pylori
PT   infection associated disease
XX
PS   Claim 2; Fig 3A-R; 303pp; English.
XX
CC   The present sequence is the complete nucleotide sequence of the CagI
CC   locus, including the true 5'-terminus of CagI. The sequence was
CC   constructed using overlapping clones (one of which also overlaps with
CC   the CagA region). The CagI region contains clusters of putative open
CC   reading frames (ORFs) with different polarities. The putative ORFs for
CC   this region are shown above. It is hypothesised that some of these ORFs
CC   may encode exporter molecules with homology to the ptl genes of
CC   Bordetella pertussis and VIR B4 genes of Agrobacterium tumefaciens and
CC   for proteins with motifs shared by the purported invasion factors of
CC   Salmonella genus. The absence of the CagA gene in the type I strains is
CC   associated with the absence of CagI sequences (which may encode virulence
CC   factors restricted to type I strains). The CagI nucleotide sequence, its
```



```
CC   fragments and encoded proteins are used in the diagnosis of H. pylori
CC   (esp. H. pylori type I strain) infection in an individual and in vaccines
CC   (claimed) for the treatment of H. pylori infection associated with e.g.
CC   duodenal and gastric ulcers, severe forms of active gastritis (esp. type
CC   gastritis) and gastric adenocarcinoma.
XX
SQ   Sequence 19932 BP; 6677 A; 3635 C; 3483 G; 6137 T; 0 other;

  Query Match            100.0%;  Score 5925;  DB 17;  Length 19932;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 5925;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctccatttttaagcaactccatagaccactaaagaaaccttttttgaggctatctttgaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14008 ctccatttttaagcaactccatagaccactaaagaaaccttttttgaggctatctttgaaa 14067

Qy      61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14068 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaac 14127

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14128 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 14187

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14188 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 14247

Qy     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcattttagcaaatttt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14248 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcattttagcaaatttt 14307

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14308 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 14367

Qy     361 ctttttttcaatattaataatgattaatgaaaaaaaaaaaaaaaatgcttgatattgttgtat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14368 ctttttttcaatattaataatgattaatgaaaaaaaaaaaaaaaatgcttgatattgttgtat 14427

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14428 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 14487

Qy     481 ttgacactaacaagataccgataggtatgaaactaggtatagtaaggagaaacaatgact 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14488 ttgacactaacaagataccgataggtatgaaactaggtatagtaaggagaaacaatgact 14547

Qy     541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   14548 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 14607

Qy     601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14608 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 14667

Qy     661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14668 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 14727

Qy     721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14728 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 14787

Qy     781 ttttcagacttttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14788 ttttcagacttttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 14847

Qy     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14848 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 14907

Qy     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14908 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatg 14967

Qy     961 gaaaatatcatacaacccccctatccttgatgataaagagaaagcggagttttttgaaatct 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14968 gaaaatatcatacaacccccctatccttgatgataaagagaaagcggagttttttgaaatct 15027

Qy    1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15028 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 15087

Qy    1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15088 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 15147

Qy    1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15148 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 15207

Qy    1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15208 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 15267

Qy    1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15268 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 15327

Qy    1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15328 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 15387

Qy    1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15388 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 15447

Qy    1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15448 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 15507

Qy    1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15508 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 15567

Qy    1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15568 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 15627

Qy    1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15628 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 15687

Qy    1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaaatagatttcatggaattt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15688 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaaatagatttcatggaattt 15747

Qy    1741 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15748 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 15807

Qy    1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15808 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 15867

Qy    1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15868 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 15927

Qy    1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15928 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 15987

Qy    1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15988 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 16047

Qy    2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16048 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 16107
```

```
Qy   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16108 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 16167

Qy   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16168 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 16227

Qy   2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttttgagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16228 aaaggattgtccccacaagaagctaataagcttatcaaagatttttttgagcagcaacaaa 16287

Qy   2281 gaattggttggaaaaacttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16288 gaattggttggaaaaacttaaacttcaataaagctgtagctgacgctaaaaacacaggc 16347

Qy   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16348 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 16407

Qy   2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16408 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 16467

Qy   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16468 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 16527

Qy   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16528 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 16587

Qy   2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16588 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 16647

Qy   2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagccctt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16648 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagccctt 16707

Qy   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaaccctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16708 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaaccctt 16767

Qy   2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16768 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 16827

Qy   2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16828 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 16887
```



```
Qy   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16888 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 16947

Qy   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaaggagcaattggcccaacaagct 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16948 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaacaagct 17007

Qy   3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17008 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 17067

Qy   3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17068 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 17127

Qy   3121 aaaaacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17128 aaaaacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 17187

Qy   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17188 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 17247

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17248 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 17307

Qy   3301 attgaccgactcaatcaaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17308 attgaccgactcaatcaaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 17367

Qy   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17368 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 17427

Qy   3421 gaatttggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17428 gaatttggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 17487

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17488 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 17547

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17548 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 17607

Qy   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db  17608 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 17667

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17668 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 17727

Qy   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17728 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 17787

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17788 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 17847

Qy   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17848 gtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctacagcatct 17907

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17908 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 17967

Qy   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccaccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17968 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccaccccttgcta 18027

Qy   4021 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18028 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttggga 18087

Qy   4081 tgaatgctaccaattcatggtatcatatcccctacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18088 tgaatgctaccaattcatggtatcatatcccctacattcgtatctagcgtaggaagtgt 18147

Qy   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18148 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 18207

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18208 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 18267

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18268 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 18327

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18328 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 18387

Qy   4381 aagacttaacccttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18388 aagacttaacccttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 18447

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18448 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 18507

Qy   4501 aagatcaagttcaaaaaatcatagagcttttagagcaaattgatcgcgctcttaacccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18508 aagatcaagttcaaaaaatcatagagcttttagagcaaattgatcgcgctcttaacccaaa 18567

Qy   4561 gaaaaatcagaaaaaccataggaattatcacacctttataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18568 gaaaaatcagaaaaaccataggaattatcacacctttataatgcccaaaaaagacgcttgc 18627

Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18628 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 18687

Qy   4681 atgccttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18688 atgccttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 18747

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18748 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 18807

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18808 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 18867

Qy   4861 tctttagcgctattttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18868 tctttagcgctattttgcaagtctgtagataggtaatcttttccaaagataatcattaga 18927

Qy   4921 cattcttcgcttcaaaacgctttcataaaatctctctcaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18928 cattcttcgcttcaaaacgctttcataaaatctctctcaaagcgctttataatcaacacaat 18987

Qy   4981 acccttatagtgtgagctatagccccttttttgggaattgagttattttgacttttaaattt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18988 acccttatagtgtgagctatagccccttttttgggaattgagttattttgacttttaaattt 19047

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19048 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 19107

Qy   5101 gctcgcatgaaattccacttttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19108 gctcgcatgaaattccacttttagggaatgcgtgtgcatttttttttaagggcgtatttttg 19167
```

```
Qy    5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggtgttgaaagggc 5220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19168 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggtgttgaaagggc 19227

Qy    5221 aaaatgctccatagaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaaccc 5280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19228 aaaatgctccatagaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaaccc 19287

Qy    5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19288 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 19347

Qy    5341 aacaattcgcccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19348 aacaattcgcccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 19407

Qy    5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19408 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 19467

Qy    5461 gtcgcttttgtccctagcactaaaataggggcgtttttatcttttacttgtcgcttgatc 5520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19468 gtcgcttttgtccctagcactaaaatagggggcgtttttatcttttacttgtcgcttgatc 19527

Qy    5521 gctaaaatgcttggctcaatcacgcccacaatagggatttttggaatgcttttgcatctct 5580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19528 gctaaaatgcttggctcaatcacgcccacaatagggatttttggaatgcttttgcatctct 19587

Qy    5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19588 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 19647

Qy    5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19648 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 19707

Qy    5701 taaggcactctagccgtatcgccatataagatgatttcatcaaataattgcgcttttaaa 5760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19708 taaggcactctagccgtatcgccatataagatgatttcatcaaataattgcgcttttaaa 19767

Qy    5761 aggctttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctattttcatgaca 5820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19768 aggctttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctattttcatgaca 19827

Qy    5821 ctttttaatttaatgggattaattagggatttttatttttcattcattaagtttaaaaat 5880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19828 ctttttaatttaatgggattaattagggatttttatttttcattcattaagtttaaaaat 19887

Qy    5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
           |||||||||||||||||||||||||||||||||||||||||||
Db   19888 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 19932
```



```
RESULT   3
V62461
ID   V62461 standard; DNA; 4149 BP.
XX
AC   V62461;
XX
DT   15-FEB-1999  (first entry)
XX
DE   Helicobacter pylori cagA-Vibrio cholerae A2+B subunit gene fusion.
XX
KW   Vaccine; antigen; CagB; antigen; toxin; diagnosis; gastritis;
KW   ulcer; stomach cancer; as.
XX
OS   Chimeric - Helicobacter pylori.
OS   Chimeric - Vibrio cholerae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3444
FT                   /*tag= a
FT                   /product= CagA
FT   CDS             3444..4020
FT                   /*tag= b
FT                   /product= A2+B subunits
XX
PN   WO9844130-A1.
XX
PD   08-OCT-1998.
XX
PF   31-MAR-1998;   98WO-KR00073.
XX
PR   31-MAR-1997;   97KR-0011951.
PR   31-MAR-1997;   97KR-0011950.
XX
PA   (DAEW-) DAEWOONG PHARM CO LTD.
XX
PI   Choi D,  Jung H,  Kim B,  Park M,  Shin S,  Yu Y;
XX
DR   WPI; 1998-568279/48.
DR   P-PSDB; W80600.
XX
PT   New chimeric proteins for use against Helicobacter pylori
PT   comprising an antigenic protein of H. pylori and A1 and B subunits
PT   of Vibrio cholerae toxin, preferably produced by recombinant
PT   techniques
XX
PS   Claim 4; Page 78-80; 102pp; English.
XX
CC   This is the nucleotide sequence of a fusion gene prepared by
CC   ligating the cagA gene of Helicobacter pylori and the A2 and B
```



```
CC   subunits genes of Vibrio cholerae toxin.  It codes for a claimed
CC   1338-amino acid fusion protein (see W80600).  The invention relates
CC   to: chimeric proteins comprising antigenic proteins of H. pylori
CC   and A2 and B subunits of V. cholerae toxin, recombinant DNAs
CC   encoding such chimeric proteins; recombinant expression vectors;
CC   a process for preparing the chimeric proteins using transformed
CC   microbial host cells; and to preventative and therapeutic vaccines
CC   comprising the chimeric proteins for H. pylori-associated diseases
CC   such as gastritis, gastric ulcer, duodenal ulcer and gastric cancer
CC   (claimed).  The chimeric proteins are designed to have excellent
CC   immunogenicity, to be stable in the stomach, to penetrate the
CC   mucous membrane of the intestine, and to stimulate production of
CC   sIgA.  They can additionally be used as active ingredients in
CC   diagnostic kits for H. pylori infection, and for production of
CC   anti-H. pylori antibody.
XX
SQ   Sequence 4149 BP; 1544 A; 714 C; 819 G; 1072 T; 0 other;

  Query Match           58.1%;  Score 3444;  DB 19;  Length 4149;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 3444;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     535 atgactaacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaa 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atgactaacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaa 60

Qy     595 tttatcaataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgat 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tttatcaataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgat 120

Qy     655 cctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaagga 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 cctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaagga 180

Qy     715 atctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaat 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 atctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaat 240

Qy     775 cagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcatt 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 cagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcatt 300

Qy     835 gatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttc 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttc 360

Qy     895 acaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaat 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 acaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaat 420
```



```
Qy     955 tttatggaaaatatcatacaaccccctatccttgatgataaagagaaagcggagtttttg 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tttatggaaaatatcatacaaccccctatccttgatgataaagagaaagcggagtttttg 480

Qy    1015 aaatctgccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaa 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 aaatctgccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaa 540

Qy    1075 aagttcatgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatgga 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 aagttcatgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatgga 600

Qy    1135 gagcctactggtggggattggttggatatttttctctcatttatatttgacaaaaaacaa 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 gagcctactggtggggattggttggatatttttctctcatttatatttgacaaaaaacaa 660

Qy    1195 tcttctgatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatata 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 tcttctgatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatata 720

Qy    1255 gccactaccaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaa 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 gccactaccaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaa 780

Qy    1315 aggggtaatttttctaaattcactcttggcgatatggaaatgttagatgttgaggagtc 1374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 aggggtaatttttctaaattcactcttggcgatatggaaatgttagatgttgaggagtc 840

Qy    1375 gctgacattgatcccaattacaagttcaatcaattattgattcacaataacgctctgtct 1434
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 gctgacattgatcccaattacaagttcaatcaattattgattcacaataacgctctgtct 900

Qy    1435 tctgtgttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatggg 1494
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 tctgtgttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatggg 960

Qy    1495 ggcaatggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaa 1554
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ggcaatggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaa 1020

Qy    1555 caaggcaacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtc 1614
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 caaggcaacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtc 1080

Qy    1615 atagcaggtggtgagaaagggattaacaaccctagttttatctctacaaagaagaccaa 1674
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 atagcaggtggtgagaaagggattaacaaccctagttttatctctacaaagaagaccaa 1140

Qy    1675 ctcacaggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatg 1734
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ctcacaggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatg 1200
```

```
Qy   1735 gaattttcttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaa 1794
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 gaatttcttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaa 1260

Qy   1795 ttccgaactgagattaaagatttccaaaaagactctaaggcttatttagacgcccctaggg 1854
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ttccgaactgagattaaagatttccaaaaagactctaaggcttatttagacgcccctaggg 1320

Qy   1855 aatgatcgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgag 1914
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 aatgatcgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgag 1380

Qy   1915 tttggtaatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagct 1974
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 tttggtaatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagct 1440

Qy   1975 ttagatagggagaaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgttt 2034
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ttagatagggagaaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgttt 1500

Qy   2035 gttgattattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggc 2094
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 gttgattattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggc 1560

Qy   2095 gttacgaatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttg 2154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 gttacgaatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttg 1620

Qy   2155 cctgatttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaacta 2214
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 cctgatttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaacta 1680

Qy   2215 accactaaaggattgtccccacaagaagctaataagcttatcaaagatttttttgagcagc 2274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 accactaaaggattgtccccacaagaagctaataagcttatcaaagatttttttgagcagc 1740

Qy   2275 aacaaagaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaac 2334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 aacaaagaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaac 1800

Qy   2335 acaggcaatttatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaa 2394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 acaggcaatttatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaa 1860

Qy   2395 cgagagctatttagagaaagaagtagagaaaaaaattggagagcaaaagcggcaacaaaaat 2454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 cgagagctatttagagaaagaagtagagaaaaaaattggagagcaaaagcggcaacaaaaat 1920

Qy   2455 aaaatggaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaat 2514
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   1921 aaaatggaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaat 1980

Qy   2515 aaagaggctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaa 2574
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 aaagaggctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaa 2040

Qy   2575 agggaattgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatct 2634
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 agggaattgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatct 2100

Qy   2635 tttgatgaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaa 2694
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 tttgatgaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaa 2160

Qy   2695 gcccttaaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaa 2754
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 gcccttaaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaa 2220

Qy   2755 aaccttaatgcagctttgaatgaatttcaaaaatggcaaaaataaggatttcagcaaggta 2814
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 aaccttaatgcagctttgaatgaatttcaaaaatggcaaaaataaggatttcagcaaggta 2280

Qy   2815 acgcaagcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggta 2874
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 acgcaagcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggta 2340

Qy   2875 acggataaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttc 2934
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 acggataaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttc 2400

Qy   2935 agtagggtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaa 2994
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 agtagggtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaa 2460

Qy   2995 caagctcaaaaaaatgaaagtctcaatgctagaaaaaaaatctgaaatatatcaatccgtt 3054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 caagctcaaaaaaatgaaagtctcaatgctagaaaaaaaatctgaaatatatcaatccgtt 2520

Qy   3055 aagaatggtgtgaatggaaccctagtcggtaatgggtatctcaagcagaagccacaact 3114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 aagaatggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaact 2580

Qy   3115 ctttctaaaaacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaat 3174
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 ctttctaaaaacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaat 2640

Qy   3175 aacaataacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaa 3234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 aacaataacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaa 2700

Qy   3235 gcagggcaagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaat 3294
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 gcagggcaagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaat 2760

Qy   3295 gcaaaaattgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcg 3354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 gcaaaaattgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcg 2820

Qy   3355 ggcttccctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaagg 3414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 ggcttccctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaagg 2880

Qy   3415 aatcaagaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagca 3474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 aatcaagaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagca 2940

Qy   3475 ggttttttttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaat 3534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 ggttttttttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaat 3000

Qy   3535 cccatgaatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaacta 3594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 cccatgaatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaacta 3060

Qy   3595 gacaattacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatc 3654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 gacaattacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatc 3120

Qy   3655 aatgaaaaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaat 3714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 aatgaaaaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaat 3180

Qy   3715 gataagatagttgcgcataatgtaggaagcgttccttgtcagagtatgataaaattggc 3774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 gataagatagttgcgcataatgtaggaagcgttccttgtcagagtatgataaaattggc 3240

Qy   3775 ttcaaccagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaac 3834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 ttcaaccagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaac 3300

Qy   3835 aatgctgtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctaca 3894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 aatgctgtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctaca 3360

Qy   3895 gcatcttattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatataca 3954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 gcatcttattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatataca 3420

Qy   3955 aaaggtggtttccaaaaatcttaa 3978
          |||||||||||||||||||||||
Db   3421 aaaggtggtttccaaaaatcttaa 3444
```



```
RESULT   4
V73601
ID   V73601 standard; DNA; 449 BP.
XX
AC   V73601;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment AMS23001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
```

```
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 166 A; 98 C; 70 G; 115 T; 0 other;

  Query Match             4.7%;  Score 277;  DB 19;  Length 449;
  Best Local Similarity   99.3%;  Pred. No. 2.4e-108;
  Matches  427;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 79

Qy    614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 139

Qy    674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    140 ttgataagaacgatagggataacaggcaagcttttgatggaatctcgcaattaagggaag 199

Qy    734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 259

Qy    794 tcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaa 853
          |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    260 tcaataagagcaatgatttaatcaacaaagacgctctcattgatgtagaatcttccacaa 319

Qy    854 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 379

Qy    914 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaaatatcatac 973
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaaatatcatac 439

Qy    974 aacccctat 983
          |||||||||
Db    440 aacccctat 449


RESULT   5
V73602
ID   V73602 standard; DNA; 449 BP.
XX
```



```
AC   V73602;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment 8832001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
```



```
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 165 A; 97 C; 71 G; 116 T; 0 other;

  Query Match             4.7%;  Score 277;  DB 19;  Length 449;
  Best Local Similarity   99.3%;  Pred. No. 2.4e-108;
  Matches  427;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 79

Qy    614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 139

Qy    674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    140 ttgataagaacgatagggataacaggcaagcttttgatggaatctcgcaattaagggaag 199

Qy    734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 259

Qy    794 tcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaa 853
          |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    260 tcaataagagcaatgatttaatcaacaaagacgctctcattgatgtagaatcttccacaa 319

Qy    854 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 379

Qy    914 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaaatatcatac 973
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaaatatcatac 439

Qy    974 aacccctat 983
          |||||||||
Db    440 aacccctat 449


RESULT   6
V73610
ID   V73610 standard; DNA; 449 BP.
XX
AC   V73610;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment I17001C2.
```



```
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 164 A; 95 C; 71 G; 119 T; 0 other;
```

```
  Query Match             4.6%;  Score 272;  DB 19;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-106;
  Matches  272;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    712 ggaatctcgcaatttaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 ggaatctcgcaatttaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 237

Qy    772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 297

Qy    832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 357

Qy    892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 417

Qy    952 aatttttatggaaaatatcatacaaccccctat 983
          |||||||||||||||||||||||||||||||||
Db    418 aatttttatggaaaatatcatacaaccccctat 449


RESULT   7
Q64580
ID   Q64580 standard; DNA; 3648 BP.
XX
AC   Q64580;
XX
DT   13-DEC-1994  (first entry)
XX
DE   120-128 kilodalton antigen of Helicobacter pylori.
XX
KW   Tag A; antigen; peptic ulceration; Helicobacter pylori; antibody;
KW   detection; diagnosis; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1072..3648
FT                   /*tag=  a
FT                   /product= 120-128 kilodalton antigen.
XX
PN   WO9409023-A.
XX
PD   28-APR-1994.
XX
PF   13-OCT-1993;   93WO-US09782.
XX
```



```
PR   13-OCT-1992;   92US-0959940.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Blaser MJ,  Cover TL,  Tummuru MKR;
XX
DR   WPI; 1994-151235/18.
DR   P-PSDB; R53268.
XX
PT   DNA coding tag A gene, from Helicobacter pylori - useful for
PT   detecting predisposition to peptic ulceration
XX
PS   Claim 2; Page 54-58; 87pp; English.
XX
CC   The antigenic fragment is a fragment of the tag A antigen of
CC   Helicobactor pylori.  Monoclonal antibodies directed against the
CC   antigen may be used to detect tag A antigen presence which is
CC   indicative of a predisposition to peptic ulceration.
XX
SQ   Sequence 3648 BP; 1298 A; 660 C; 692 G; 998 T; 0 other;


  Query Match             4.0%;  Score 235;  DB 15;  Length 3648;
  Best Local Similarity   99.7%;  Pred. No. 1.5e-90;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1087 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 1146

Qy    610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1147 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 1206

Qy    670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1207 atcgttgataagaacgatagggataacaggcaagcttttgagggaatctcgcaattaagg 1266

Qy    730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1267 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 1326

Qy    790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1327 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 1372


RESULT   8
T14052
ID   T14052 standard; DNA; 3648 BP.
XX
```



```
AC   T14052;
XX
DT   09-SEP-1996  (first entry)
XX
DE   Helicobacter pylori TagA DNA encoding a 96 kD antigenic fragment.
XX
KW   Antigen; peptic ulcer; chronic gastritis; gastric adenocarcinoma;
KW   diagnosis: predisposition; antibody; vaccine; infection; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1072..3648
FT                   /*tag=  a
FT   5'UTR           1..1071
FT                   /*tag=  b
XX
PN   WO9610639-A2.
XX
PD   11-APR-1996.
XX
PF   29-SEP-1995;   95WO-US12669.
XX
PR   30-SEP-1994;   94US-0316397.
XX
PA   (ORAV-) ORAVAX INC.
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Blaser MJ,  Cover TL,  Kleanthous H,  Tummuru MKR;
XX
DR   WPI; 1996-209361/21.
DR   P-PSDB; R91308.
XX
PT   Helicobacter pylori Tag A gene - used to develop prods. for the
PT   diagnosis, treatment and prevention of peptic ulceration and gastric
PT   carcinoma
XX
PS   Claim 5; Page 88-93; 118pp; English.
XX
CC   T14052 encodes a 96 kD antigenic fragment of the Helicobacter pylori
CC   (HP) TagA antigen. TagA, or its fragments, can be used to determine
CC   a predisposition to peptic ulceration or gastric carcinoma, both
CC   conditions caused by or linked to HP infection. TagA and its
CC   fragments may also be used for antibody prodn. for use in detection
CC   of TagA in patients suspected of HP infection. TagA antibodies or
CC   other ligands may also be used to treat peptic ulceration or gastric
CC   carcinoma caused by HP infection. TagA or a non-functional TagA
CC   mutant may be used in vaccines for preventing and treating HP
CC   infection.
XX
SQ   Sequence 3648 BP; 1298 A; 660 C; 692 G; 998 T; 0 other;
```



```
  Query Match             4.0%;  Score 235;  DB 17;  Length 3648;
  Best Local Similarity   99.7%;  Pred. No. 1.5e-90;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1087 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 1146

Qy    610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1147 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 1206

Qy    670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1207 atcgttgataagaacgatagggataacaggcaagcttttgagggaatctcgcaattaagg 1266

Qy    730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1267 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 1326

Qy    790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1327 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 1372


RESULT   9
Q64581
ID   Q64581 standard; DNA; 4821 BP.
XX
AC   Q64581;
XX
DT   13-DEC-1994  (first entry)
XX
DE   Tag A antigen of Helicobacter pylori.
XX
KW   Tag A; antigen; peptic ulceration; Helicobacter pylori; antibody;
KW   detection; diagnosis; treatment; therapy; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1072..4614
FT                   /*tag=  a
FT                   /product= Tag A antigen of Helicobacter pylori.
XX
PN   WO9409023-A.
XX
PD   28-APR-1994.
XX
```

```
PF   13-OCT-1993;   93WO-US09782.
XX
PR   13-OCT-1992;   92US-0959940.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Blaser MJ,  Cover TL,  Tummuru MKR;
XX
DR   WPI; 1994-151235/18.
DR   P-PSDB; R53269.
XX
PT   DNA coding tag A gene, from Helicobacter pylori - useful for
PT   detecting predisposition to peptic ulceration
XX
PS   Claim 27; Page 63-68; 87pp; English.
XX
CC   Monoclonal antibodies directed against the antigen may be used to
CC   detect tag A antigen presence which is indicative of a predisposition
CC   to peptic ulceration.  A ligand e.g. Antibody, specifically reactive
CC   with the tag A antigen can be used to treat peptic ulcers.
XX
SQ   Sequence 4821 BP; 1722 A; 874 C; 929 G; 1296 T; 0 other;


  Query Match             4.0%;  Score 235;  DB 15;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 1.5e-90;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1207 atcgttgataagaacgatagggataacaggcaagcttttgagggaatctcgcaattaagg 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 1372


RESULT  10
Q86728
```



```
ID   Q86728 standard; DNA; 4821 BP.
XX
AC   Q86728;
XX
DT   29-SEP-1995  (first entry)
XX
DE   H. pylori tagA antigen-encoding DNA.
XX
KW   TagA; antigen; ulcer; diagnosis; vaccine; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1072..4617
FT                   /*tag=  a
XX
PN   US5403924-A.
XX
PD   04-APR-1995.
XX
PF   13-OCT-1992;   92US-0959940.
XX
PR   13-OCT-1992;   92US-0959940.
PR   26-APR-1993;   93US-0053614.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Blaser MJ,  Cover TL,  Tummuru MKR;
XX
DR   WPI; 1995-146855/19.
DR   P-PSDB; R72593.
XX
PT   New nucleic acid encoding tag A antigen of Helicobacter pylori  -
PT   used to detect predispositon to peptic ulceration and to produce
PT   protein for use in vaccines, diagnosis etc.
XX
PS   Disclosure; Column 37-46; 30pp; English.
XX
CC   The full-length sequence of the tagA gene of H. pylori 84-183 (ATCC
CC   53726) was obtained from overlapping clones isolated from genomic
CC   libraries.  Vectors carrying the gene are used for production
CC   of recombinant tagA antigen in host cells.
XX
SQ   Sequence 4821 BP; 1721 A; 874 C; 930 G; 1296 T; 0 other;


  Query Match             4.0%;  Score 235;  DB 16;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 1.5e-90;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
```



```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1207 atcgttgataagaacgatagggataacaggcaagcttttgagggaatctcgcaattaagg 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 1372


RESULT  11
T14051
ID   T14051 standard; DNA; 4821 BP.
XX
AC   T14051;
XX
DT   09-SEP-1996  (first entry)
XX
DE   Helicobacter pylori TagA gene encoding a 120-128 kD antigen.
XX
KW   Antigen; peptic ulcer; chronic gastritis; gastric adenocarcinoma;
KW   diagnosis; predisposition; antibody; vaccine; infection; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1072..4617
FT                   /*tag=  a
FT   5'UTR           1..1071
FT                   /*tag=  b
FT   3'UTR           4618..4821
FT                   /*tag=  c
XX
PN   WO9610639-A2.
XX
PD   11-APR-1996.
XX
PF   29-SEP-1995;   95WO-US12669.
XX
PR   30-SEP-1994;   94US-0316397.
XX
```



```
PA   (ORAV-) ORAVAX INC.
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Blaser MJ,  Cover TL,  Kleanthous H,  Tummuru MKR;
XX
DR   WPI; 1996-209361/21.
DR   P-PSDB; R91307.
XX
PT   Helicobacter pylori Tag A gene - used to develop prods. for the
PT   diagnosis, treatment and prevention of peptic ulceration and gastric
PT   carcinoma
XX
PS   Claim 4; Page 96-102; 118pp; English.
XX
CC   T14051 encodes a 120-128 kD antigen of Helicobacter pylori (HP). The
CC   antigen is designated TagA and it, or its fragments, can be used to
CC   determine a predisposition to peptic ulceration or gastric carcinoma,
CC   conditions caused by or linked to HP infection. TagA and its
CC   fragments may also be used for antibody prodn. for use in detection
CC   of TagA in patients suapected of HP infection. TagA antibodies or
CC   other ligands may also be used to treat peptic ulceration or gastric
CC   carcinoma caused by HP infection. TagA or a non-functional TagA
CC   mutant may be used in vaccines for preventing and treating HP
CC   infection.
XX
SQ   Sequence 4821 BP; 1721 A; 875 C; 929 G; 1296 T; 0 other;


  Query Match             4.0%;  Score 235;  DB 17;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 1.5e-90;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1207 atcgttgataagaacgatagggataacaggcaagcttttgagggaatctcgcaattaagg 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 1372
```

RESULT 12
V73600
ID V73600 standard; DNA; 449 BP.
XX
AC V73600;
XX
DT 22-MAR-2000 (first entry)
XX
DE H. pylori cagA DNA fragment AMST7001C2.
XX
KW PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS Helicobacter pylori.
XX
PN WO9816658-A2.
XX
PD 23-APR-1998.
XX
PF 10-OCT-1997; 97WO-EP05614.
XX
PR 16-OCT-1996; 96EP-0870131.
PR 09-SEP-1997; 97EP-0870133.
XX
PA (INNO-) INNOGENETICS NV.
PA (DDLD-) DDL BV.
XX
PI Quint W, Van Doorn L;
XX
DR WPI; 1998-251300/22.
XX
PT Method for detecting and/or typing Helicobacter pylori strains -
PT comprises use of primers and probes based on vacA and cagA gene
XX
PS Claim 18; Figure 10; 122pp; English.
XX
CC This invention describes a novel method for the detection and/or typing
CC of Helicobacter pylori strains present in a sample using PCR primers and
CC probes to detect regions of the vacuolating toxin (vacA) gene and other
CC virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC gene. The method allows the typing and allele-specific detection of a
CC strain according to the VDG alleles present in that particular H. pylori
CC strain. The virulence determinant genes are the genetic elements involved
CC in enabling, determining, and marking the infectivity and/or
CC pathogenicity of the H. pylori strain. The method provides a way of
CC detecting H. pylori strains in a sample with respect to the development

CC of chronic active gastritis, gastric and duodenal ulcers, gastric
CC adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC determining eradication therapy. V73508-V73546 represent PCR primers
CC and probes used in the detection of the H. pylori vacA and cagA genes.
CC The primers and probes are used especially to detect the vacA S regions
CC S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC V73547-V73785.
XX
SQ Sequence 449 BP; 166 A; 96 C; 71 G; 116 T; 0 other;

Query Match 3.5%; Score 210; DB 19; Length 449;
Best Local Similarity 99.0%; Pred. No. 7.4e-80;
Matches 410; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 570 cgaagcggctttaacccgcagcaatttatcaataatcttcaagtagcttttcttaaagt 629
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 36 cgaagcggctttaacccgcagcaatttatcaataatcttcaagtagcttttcttaaagt 95

Qy 630 tgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcgttgataagaacgatag 689
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 96 tgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcgttgataagaacgatag 155

Qy 690 ggataacaggcaagcttttgaaggaatctcgcaatttaagggaagaatactccaataaagc 749
|||||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db 156 ggataataggcaagcttttgatggaatctcgcaatttaagggaagaatactccaataaagc 215

Qy 750 gatcaaaaatcctaccaaaaaagaatcagtatttttcagactttatcaataagagcaatga 809
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 216 gatcaaaaatcctaccaaaaaagaatcagtatttttcagactttatcaataagagcaatga 275

Qy 810 tttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaatt 869
||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db 276 tttaatcaacaaagacgctctcattgatgtagaatcttccacaaagagctttcagaaatt 335

Qy 870 tggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaa 929
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 336 tggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaa 395

Qy 930 aatcaacacccgatcgatccgaaatttttatggaaaaatatcatacaaccccctat 983
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 396 aatcaacacccgatcgatccgaaatttttatggaaaaatatcatacaaccccctat 449

RESULT 13
V73606
ID V73606 standard; DNA; 449 BP.
XX
AC V73606;
XX
DT 22-MAR-2000 (first entry)

XX
DE H. pylori cagA DNA fragment AUS5001C2.
XX
KW PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS Helicobacter pylori.
XX
PN WO9816658-A2.
XX
PD 23-APR-1998.
XX
PF 10-OCT-1997; 97WO-EP05614.
XX
PR 16-OCT-1996; 96EP-0870131.
PR 09-SEP-1997; 97EP-0870133.
XX
PA (INNO-) INNOGENETICS NV.
PA (DDLD-) DDL BV.
XX
PI Quint W, Van Doorn L;
XX
DR WPI; 1998-251300/22.
XX
PT Method for detecting and/or typing Helicobacter pylori strains -
PT comprises use of primers and probes based on vacA and cagA gene
XX
PS Claim 18; Figure 10; 122pp; English.
XX
CC This invention describes a novel method for the detection and/or typing
CC of Helicobacter pylori strains present in a sample using PCR primers and
CC probes to detect regions of the vacuolating toxin (vacA) gene and other
CC virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC gene. The method allows the typing and allele-specific detection of a
CC strain according to the VDG alleles present in that particular H. pylori
CC strain. The virulence determinant genes are the genetic elements involved
CC in enabling, determining, and marking the infectivity and/or
CC pathogenicity of the H. pylori strain. The method provides a way of
CC detecting H. pylori strains in a sample with respect to the development
CC of chronic active gastritis, gastric and duodenal ulcers, gastric
CC adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC determining eradication therapy. V73508-V73546 represent PCR primers
CC and probes used in the detection of the H. pylori vacA and cagA genes.
CC The primers and probes are used especially to detect the vacA S regions
CC S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC V73547-V73785.
XX
SQ Sequence 449 BP; 168 A; 95 C; 69 G; 117 T; 0 other;

Query Match 3.3%; Score 194; DB 19; Length 449;
Best Local Similarity 99.6%; Pred. No. 4.7e-73;
Matches 244; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 739 tccaataaagcgatcaaaaatcctaccaaaaaagaatcagtatttttcagactttatcaat 798
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 205 tccaataaagcgatcaaaaatcctaccaaaaaagaatcagtatttttcagactttatcgat 264

Qy 799 aagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagc 858
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 265 aagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagc 324

Qy 859 tttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaac 918
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 325 tttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaac 384

Qy 919 gatccgtctaaaatcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccc 978
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 385 gatccgtctaaaatcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccc 444

Qy 979 cctat 983
|||||
Db 445 cctat 449

RESULT 14
V73609
ID V73609 standard; DNA; 449 BP.
XX
AC V73609;
XX
DT 22-MAR-2000 (first entry)
XX
DE H. pylori cagA DNA fragment AI9001C2.
XX
KW PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS Helicobacter pylori.
XX
PN WO9816658-A2.
XX
PD 23-APR-1998.
XX
PF 10-OCT-1997; 97WO-EP05614.
XX

```
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 162 A; 96 C; 72 G; 119 T; 0 other;


  Query Match             3.0%;  Score 179;  DB 19;  Length 449;
  Best Local Similarity   99.1%;  Pred. No. 1.1e-66;
  Matches  329;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     652 gatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaa 711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 gatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgat 177

Qy     712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 237

Qy     772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
           ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db     238 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacgctctc 297
```



```
Qy     832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 357

Qy     892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 951
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 417

Qy     952 aattttatggaaaatatcatacaaccccctat 983
           ||||||||||||||||||||||||||||||||
Db     418 aattttatggaaaatatcatacaaccccctat 449


RESULT  15
V73611
ID   V73611 standard; DNA; 449 BP.
XX
AC   V73611;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment AM82001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
```



```
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 164 A; 96 C; 71 G; 118 T; 0 other;


  Query Match             3.0%;  Score 179;  DB 19;  Length 449;
  Best Local Similarity   99.1%;  Pred. No. 1.1e-66;
  Matches  329;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     652 gatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaa 711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 gatcctgatcaaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgat 177

Qy     712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 237

Qy     772 aatcagtatttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctc 831
           ||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db     238 aatcagtatttttcagactttatcgataagagcaacgatttaatcaacaaagacaatctc 297

Qy     832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 357

Qy     892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 951
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccga 417

Qy     952 aattttatggaaaatatcatacaaccccctat 983
           ||||||||||||||||||||||||||||||||
Db     418 aattttatggaaaatatcatacaaccccctat 449
```



```
RESULT  16
V73604
ID   V73604 standard; DNA; 449 BP.
XX
AC   V73604;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment AM545001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
```

```
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 169 A; 97 C; 69 G; 114 T; 0 other;


  Query Match             3.0%;  Score 175;  DB 19;  Length 449;
  Best Local Similarity   98.8%;  Pred. No. 5.6e-65;
  Matches  425;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     554 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 79

Qy     614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 tggcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 139

Qy     674 ttgataagaacgatagggataacaggcaagcttttgaagggaatctcgcaattaagggaag 733
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     140 ttgataagaacgatagggataacaggcaagcttttgatggaatctcgcaattaagggaag 199

Qy     734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     200 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcaaacttta 259

Qy     794 tcaataagagcaatgatttaatcaacaaagacaatctcattgatgtagaatcttccacaa 853
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||| ||
Db     260 tcaataagagcaatgatctaatcaacaaagacaatctcattgatgtagaatcttccaaaa 319

Qy     854 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 agagctttcagaaatttggggatcagcgttaccgaattttcacaagttgggtgtcccatc 379

Qy     914 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaaatatcatac 973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     380 aaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatggaaaatatcatac 439

Qy     974 aacccccctat 983
           ||||||||||||
Db     440 aacccccctat 449


RESULT  17
T67410
ID   T67410 standard; DNA; 3534 BP.
XX
```



```
AC   T67410;
XX
DT   08-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein ORF 16459375.aa.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3534
FT                   /*tag= a
FT                   /note= "no stop codon is given"
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US09122.
XX
PR   01-APR-1996;   96US-0630405.
PR   07-JUN-1995;   95US-0487032.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; W20174.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
XX
PS   Claim 9; Pages 156-157; 1481pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori cytoplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds, useful
CC   as potential H. pylori life cycle activators or inhibitors.  The genomic
CC   sequence of H. pylori (ATCC 55679) was determined from overlapping
CC   contigs generated by mechanically shearing the bacterial DNA. The
CC   sequences were analysed for ORF of at least 180 nucleotides, and the
CC   predicted coding regions defined by computer evaluation. To identify
CC   likely H. pylori antigens for vaccine development, the amino acid
CC   sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
```



```
CC   isolated from H. pylori by PCR amplification for recombinant
CC   polypeptide production, e.g. in E. coli hosts.
XX
SQ   Sequence 3534 BP; 1308 A; 603 C; 721 G; 902 T; 0 other;


  Query Match             2.8%;  Score 166;  DB 18;  Length 3534;
  Best Local Similarity   99.5%;  Pred. No. 3.2e-61;
  Matches  216;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    2628 taaatcttttgatgaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaac 2687
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2139 taaatcttttgatgaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaac 2198

Qy    2688 actaaaagcccttaaagttcggtgaaagatttaggtatcaatccagaatggatttcaaa 2747
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    2199 actaaaagcccttaaaggctcggtgaaagatttaggtatcaatccagaatggatttcaaa 2258

Qy    2748 agttgaaaaccttaatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcag 2807
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2259 agttgaaaaccttaatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcag 2318

Qy    2808 caaggtaacgcaagcaaaaagcgaccttgaaaattcc 2844
           |||||||||||||||||||||||||||||||||||||
Db    2319 caaggtaacgcaagcaaaaagcgaccttgaaaattcc 2355


RESULT  18
T68137
ID   T68137 standard; DNA; 3549 BP.
XX
AC   T68137;
XX
DT   17-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein ORF 13eel2016orf74.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3549
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   WO9640893-A1.
XX
```



```
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US09122.
XX
PR   01-APR-1996;   96US-0630405.
PR   07-JUN-1995;   95US-0487032.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; W20884.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
XX
PS   Claim 9; Page 921-22; 1481pp; English.
XX
CC   This sequence encodes a H. pylori cytoplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
XX
SQ   Sequence 3549 BP; 1312 A; 604 C; 723 G; 910 T; 0 other;


  Query Match             2.8%;  Score 166;  DB 18;  Length 3549;
  Best Local Similarity   99.5%;  Pred. No. 3.2e-61;
  Matches  216;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    2628 taaatcttttgatgaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaac 2687
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2154 taaatcttttgatgaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaac 2213

Qy    2688 actaaaagcccttaaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaa 2747
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    2214 actaaaagcccttaaaggctcggtgaaagatttaggtatcaatccagaatggatttcaaa 2273

Qy    2748 agttgaaaaccttaatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcag 2807
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2274 agttgaaaacctttaatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcag 2333

Qy    2808 caaggtaacgcaagcaaaaagcgaccttgaaaattcc 2844
           |||||||||||||||||||||||||||||||||||||
Db    2334 caaggtaacgcaagcaaaaagcgaccttgaaaattcc 2370


RESULT  19
V73605
ID   V73605 standard; DNA; 449 BP.
XX
AC   V73605;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment AM79001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
```



```
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 166 A; 100 C; 70 G; 113 T; 0 other;


  Query Match             2.6%;  Score 156;  DB 19;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 6.7e-57;
  Matches  156;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     828 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     294 tctcattgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccg 353

Qy     888 aatttttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 947
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     354 aatttttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgat 413

Qy     948 ccgaaatttttatggaaaatatcatacaaccccctat 983
           |||||||||||||||||||||||||||||||||||||
Db     414 ccgaaatttttatggaaaatatcatacaaccccctat 449


RESULT  20
V73608
ID   V73608 standard; DNA; 449 BP.
XX
AC   V73608;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment I16001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
```



```
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 163 A; 95 C; 71 G; 120 T; 0 other;


  Query Match             2.3%;  Score 136;  DB 19;  Length 449;
  Best Local Similarity   99.2%;  Pred. No. 2.1e-48;
  Matches  236;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     712 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 ggaatctcgcaattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaag 237
```



```
Qy     772 aatcagtatttttcagacttttatcaataagagcaatgatttaatcaacaaagacaatctc 831
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db     238 aatcagtatttttcagacttttatcaataagagcaatgatttaatcaacaaagacgctctc 297

Qy     832 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 891
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 attgatgtagaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatt 357

Qy     892 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatcc 949
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 ttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatcc 415


RESULT  21
V73612
ID   V73612 standard; DNA; 449 BP.
XX
AC   V73612;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment AUS1001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
```

```
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomaa, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 167 A; 96 C; 68 G; 118 T; 0 other;


  Query Match             2.2%;  Score 131;  DB 19;  Length 449;
  Best Local Similarity   98.9%;  Pred. No. 2.8e-46;
  Matches  281;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggcttttaaccccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 accaacaaccacaaaccgaagcggcttttaaccccgcagcaatttatcaataatcttcaag 79

Qy    614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 tggcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 139

Qy    674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagggaag 733
          |||||||||| ||||||||||||||||||||||||| ||||||||||||||||||||||
Db    140 ttgataagaatgatagggataacaggcaagcttttgatggaatctcgcaattaagggaag 199

Qy    734 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 aatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagacttta 259

Qy    794 tcaataagagcaatgatttaatcaacaaagacaatctcattgat 837
          |||||||||||||||||||||||||||||||||||||||||||
Db    260 tcaataagagcaatgatttaatcaacaaagacaatctcattgat 303


RESULT  22
V73607
ID   V73607 standard; DNA; 449 BP.
XX
```



```
AC   V73607;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment AUS21RCAG2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains  -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
```



```
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 169 A; 95 C; 68 G; 117 T; 0 other;


  Query Match             2.1%;  Score 125;  DB 19;  Length 449;
  Best Local Similarity   99.4%;  Pred. No. 1e-43;
  Matches  175;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    808 gatttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaa 867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    274 gatttaatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaa 333

Qy    868 tttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtct 927
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    334 tttggggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaatgatccgtct 393

Qy    928 aaaatcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccccctat 983
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    394 aaaatcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccccctat 449


RESULT  23
V73613
ID   V73613 standard; DNA; 449 BP.
XX
AC   V73613;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment AM105001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
```



```
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains  -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 166 A; 94 C; 70 G; 119 T; 0 other;


  Query Match             2.0%;  Score 121;  DB 19;  Length 449;
  Best Local Similarity   99.4%;  Pred. No. 5e-42;
  Matches  171;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    812 taatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaatttg 871
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 taatcaacaaagacaatctcattgatgtagaatcttccacaaagagctttcagaaatttg 337

Qy    872 gggatcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaaa 931
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    338 gggatcagcgttaccaaattttcacaagttgggtgtcccatcaaaacgatccgtctaaaa 397

Qy    932 tcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccccctat 983
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    398 tcaacacccgatcgatccgaaatttttatggaaaatatcatacaaccccctat 449


RESULT  24
V73603
ID   V73603 standard; DNA; 449 BP.
```

```
XX
AC   V73603;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment AUS10001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
```



```
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 165 A; 98 C; 70 G; 116 T; 0 other;


  Query Match             2.0%;  Score 119;  DB 19;  Length 449;
  Best Local Similarity   98.9%;  Pred. No. 3.6e-41;
  Matches  269;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 79

Qy    614 tagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 673
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 tggcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaaccaatcg 139

Qy    674 ttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaatttaagggaag 733
          ||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||||||
Db    140 ttgataagaacgatagggataataggcaagcttttgatggaatctcgcaatttaagggaag 199

Qy    734 aatactccaataaagcgatcaaaaatcctaccaaaaaagaatcagtatttttcagacttta 793
          : |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 aatactccaataaagcgatcaaaaatcctaccaaaaaagaatcagtatttttcagacttta 259

Qy    794 tcaataagagcaatgatttaatcaacaaagac 825
          ||||||||||||||||||||||||||||||||
Db    260 tcaataagagcaatgatttaatcaacaaagac 291


RESULT  25
V90655/c
ID   V90655 standard; DNA; 285 BP.
XX
AC   V90655;
XX
DT   18-FEB-1999  (first entry)
XX
DE   Nucleotide sequence of clone Y178-1T3.SEQ from cluster 18.
XX
KW   Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW   peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9849314-A2.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-US08487.
```



```
XX
PR   14-OCT-1997;   97US-0061958.
PR   25-APR-1997;   97US-0045107.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Chow TP,  Fry KE,  Lim MY,  McAtee CP;
XX
DR   WPI; 1999-009433/01.
XX
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
XX
PS   Claim 27; Page 144; 402pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori antigenic protein
CC   that is characterised by immunoreactivity with H. pylori-positive
CC   antisera. The proteins are highly immunogenic and induce a long-lasting
CC   immune response that persists even after antimicriobial treatment. In
CC   antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC   highly sensitive and specific. The specification also describes 69
CC   previously unrecognised immunogenic cluster families. H. pylori antigens
CC   are used to detect H. pylori-specific antibodies, for diagnosing
CC   infection or to confirm eradication of infection, and in vaccines to
CC   protect against H. pylori infection and related diseases (gastritis,
CC   peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ   Sequence 285 BP; 76 A; 49 C; 59 G; 101 T; 0 other;


  Query Match             1.9%;  Score 110;  DB 20;  Length 285;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-37;
  Matches  110;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    876 tcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaa 935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    226 TCAGCGTTACCGAATTTTCACAAGTTGGGTGTCCCATCAAAACGATCCGTCTAAAATCAA 167

Qy    936 cacccgatcgatccgaaatttatggaaaatatcatacaaccccctatcc 985
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    166 CACCCGATCGATCCGAAATTTTATGGAAAATATCATACAACCCCCTATCC 117


RESULT  26
V90548
ID   V90548 standard; DNA; 2705 BP.
XX
AC   V90548;
XX
DT   18-FEB-1999  (first entry)
```



```
XX
DE   Nucleotide sequence from clone  C3.
XX
KW   Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW   peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9849314-A2.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-US08487.
XX
PR   14-OCT-1997;   97US-0061958.
PR   25-APR-1997;   97US-0045107.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Chow TP,  Fry KE,  Lim MY,  McAtee CP;
XX
DR   WPI; 1999-009433/01.
XX
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
XX
PS   Claim 27; Page 99-100; 402pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori antigenic protein
CC   that is characterised by immunoreactivity with H. pylori-positive
CC   antisera. The proteins are highly immunogenic and induce a long-lasting
CC   immune response that persists even after antimicriobial treatment. In
CC   antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC   highly sensitive and specific. The specification also describes 69
CC   previously unrecognised immunogenic cluster families. H. pylori antigens
CC   are used to detect H. pylori-specific antibodies, for diagnosing
CC   infection or to confirm eradication of infection, and in vaccines to
CC   protect against H. pylori infection and related diseases (gastritis,
CC   peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ   Sequence 2705 BP; 942 A; 490 C; 469 G; 804 T; 0 other;


  Query Match             1.9%;  Score 110;  DB 20;  Length 2705;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-37;
  Matches  110;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    876 tcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaa 935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2480 tcagcgttaccgaattttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaa 2539
```

```
Qy    936 cacccgatcgatccgaaatttttatggaaaatatcatacaaccccctatcc 985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2540 cacccgatcgatccgaaatttttatggaaaatatcatacaaccccctatcc 2589

RESULT  27
V90861
ID   V90861 standard; DNA; 2705 BP.
XX
AC   V90861;
XX
DT   18-FEB-1999  (first entry)
XX
DE   Nucleotide sequence of cluster 18.
XX
KW   Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW   peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9849314-A2.
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;   98WO-US08487.
XX
PR   14-OCT-1997;   97US-0061958.
PR   25-APR-1997;   97US-0045107.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Chow TP,  Fry KE,  Lim MY,  McAtee CP;
XX
DR   WPI; 1999-009433/01.
XX
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
XX
PS   Claim 21; Page 289-292; 402pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori antigenic protein
CC   that is characterised by immunoreactivity with H. pylori-positive
CC   antisera. The proteins are highly immunogenic and induce a long-lasting
CC   immune response that persists even after antimicriobial treatment. In
CC   antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC   highly sensitive and specific. The specification also describes 69
CC   previously unrecognised immunogenic cluster families. H. pylori antigens
CC   are used to detect H. pylori-specific antibodies, for diagnosing
CC   infection or to confirm eradication of infection, and in vaccines to
```



```
CC   protect against H. pylori infection and related diseases (gastritis,
CC   peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ   Sequence 2705 BP; 944 A; 489 C; 469 G; 803 T; 0 other;

  Query Match             1.9%;  Score 110;  DB 20;  Length 2705;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-37;
  Matches  110;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    876 tcagcgttaccgaatttttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaa 935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2480 tcagcgttaccgaatttttcacaagttgggtgtcccatcaaaacgatccgtctaaaatcaa 2539

Qy    936 cacccgatcgatccgaaatttttatggaaaatatcatacaaccccctatcc 985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2540 cacccgatcgatccgaaatttttatggaaaatatcatacaaccccctatcc 2589

RESULT  28
X13977/c
ID   X13977 standard; DNA; 386 BP.
XX
AC   X13977;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 1780 gene.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease; ss.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             54..347
FT                   /*tag= a
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
```



```
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   P-PSDB; W98258.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 1; Page 248-249; 2054pp; English.
XX
CC   This sequence represents a polynucleotide of the invention. It was
CC   isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC   The polypeptides can be used for preventing or treating Helicobacter
CC   infections, and gastroduodenal diseases associated with these
CC   infections, including acute, chronic, and atrophic gastritis, and peptic
CC   ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC   for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence 386 BP; 127 A; 69 C; 65 G; 125 T; 0 other;

  Query Match             1.3%;  Score 79;  DB 19;  Length 386;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-24;
  Matches   79;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4615 gcttgcgatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagaca 4674
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    355 GCTTGCGATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACA 296

Qy   4675 ctgtggatgcctttcaagg 4693
          ||||||||||||||||||
Db    295 CTGTGGATGCCTTTCAAGG 277

RESULT  29
T67571/c
ID   T67571 standard; DNA; 564 BP.
XX
AC   T67571;
XX
DT   11-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein ORF 30082267.aa.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis; ds.
XX
```



```
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..507
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US09122.
XX
PR   01-APR-1996;   96US-0630405.
PR   07-JUN-1995;   95US-0487032.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; W20384.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
XX
PS   Claim 9; Page 243; 1481pp; English.
XX
CC   The present sequence encodes a H. pylori cytoplasmic protein involved in
CC   amino acid metabolism.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
XX
SQ   Sequence 564 BP; 167 A; 123 C; 117 G; 157 T; 0 other;

  Query Match             1.1%;  Score 67;  DB 18;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-19;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy   5327 gcatgcaagtttctaacaattcgcccctctaaaatactttcttcaatcaaaggcacaaaaa 5386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 GCATGCAAGTTTCTAACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAA 230

Qy   5387 gagaagt 5393
          |||||||
Db    229 GAGAAGT 223


RESULT  30
T68133/c
ID   T68133 standard; DNA; 657 BP.
XX
AC   T68133;
XX
DT   17-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein ORF 13eel0216orf10.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..657
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;    96WO-US09122.
XX
PR   01-APR-1996;    96US-0630405.
PR   07-JUN-1995;    95US-0487032.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; W20880.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
```



```
XX
PS   Claim 9; Page 919; 1481pp; English.
XX
CC   This sequence encodes a H. pylori cytoplasmic protein involved in
CC   amino acid metabolism.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
XX
SQ   Sequence 657 BP; 194 A; 142 C; 136 G; 185 T; 0 other;


  Query Match             1.1%;  Score 67;  DB 18;  Length 657;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-19;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5327 gcatgcaagtttctaacaattcgcccctctaaaatactttcttcaatcaaaggcacaaaaa 5386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382 GCATGCAAGTTTCTAACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAA 323

Qy   5387 gagaagt 5393
          |||||||
Db    322 GAGAAGT 316


RESULT  31
V90662
ID   V90662 standard; DNA; 239 BP.
XX
AC   V90662;
XX
DT   18-FEB-1999  (first entry)
XX
DE   Nucleotide sequence of clone Y194-2T7.SEQ from cluster 18.
XX
KW   Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW   peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9849314-A2.
```



```
XX
PD   05-NOV-1998.
XX
PF   27-APR-1998;    98WO-US08487.
XX
PR   14-OCT-1997;    97US-0061958.
PR   25-APR-1997;    97US-0045107.
XX
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Chow TP,  Fry KE,  Lim MY,  McAtee CP;
XX
DR   WPI; 1999-009433/01.
XX
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
XX
PS   Claim 27; Page 146; 402pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori antigenic protein
CC   that is characterised by immunoreactivity with H. pylori-positive
CC   antisera. The proteins are highly immunogenic and induce a long-lasting
CC   immune response that persists even after antimicriobial treatment. In
CC   antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC   highly sensitive and specific. The specification also describes 69
CC   previously unrecognised immunogenic cluster families. H. pylori antigens
CC   are used to detect H. pylori-specific antibodies, for diagnosing
CC   infection or to confirm eradication of infection, and in vaccines to
CC   protect against H. pylori infection and related diseases (gastritis,
CC   peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ   Sequence 239 BP; 87 A; 48 C; 42 G; 62 T; 0 other;


  Query Match             1.0%;  Score 61;  DB 20;  Length 239;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-16;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    492 aagataccgataggtatgaaactaggtatagtaaggagaaacaatgactaacgaaaccat 551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39 aagataccgataggtatgaaactaggtatagtaaggagaaacaatgactaacgaaaccat 98

Qy    552 t 552
          |
Db     99 t 99


RESULT  32
V73614
ID   V73614 standard; DNA; 449 BP.
```



```
XX
AC   V73614;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment J123001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;    97WO-EP05614.
XX
PR   16-OCT-1996;    96EP-0870131.
PR   09-SEP-1997;    97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
```

```
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 449 BP; 166 A; 88 C; 73 G; 122 T; 0 other;


  Query Match             1.0%;  Score 61;  DB 19;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-16;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    554 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 accaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatcttcaag 79

Qy    614 t 614
          |
Db     80 t 80


RESULT  33
V73618
ID   V73618 standard; DNA; 464 BP.
XX
AC   V73618;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment HK8001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
```



```
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 464 BP; 175 A; 96 C; 70 G; 123 T; 0 other;


  Query Match             0.9%;  Score 56;  DB 19;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-14;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    594 atttatcaataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcat 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 atttatcaataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcat 130


RESULT  34
V73621
ID   V73621 standard; DNA; 464 BP.
XX
AC   V73621;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment HK9001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
```



```
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 464 BP; 175 A; 95 C; 71 G; 123 T; 0 other;


  Query Match             0.8%;  Score 50;  DB 19;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-12;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    763 accaaaaagaatcagtatttttcagactttatcaataagagcaatgattt 812
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 accaaaaagaatcagtatttttcagactttatcaataagagcaatgattt 293


RESULT  35
V73616
ID   V73616 standard; DNA; 464 BP.
XX
AC   V73616;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment HK7001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
```

```
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 464 BP; 177 A; 93 C; 69 G; 125 T; 0 other;


  Query Match             0.8%;  Score 49;  DB 19;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 2e-11;
  Matches   49;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     764 ccaaaaagaatcagtatttttcagactttatcaataagagcaatgattt 812
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 ccaaaaagaatcagtatttttcagactttatcaataagagcaatgattt 293


RESULT  36
V73617
ID   V73617 standard; DNA; 464 BP.
XX
AC   V73617;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment HK12001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
```



```
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains  -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 464 BP; 172 A; 96 C; 74 G; 122 T; 0 other;


  Query Match             0.8%;  Score 49;  DB 19;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 2e-11;
  Matches   49;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     764 ccaaaaagaatcagtatttttcagactttatcaataagagcaatgattt 812
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 ccaaaaagaatcagtatttttcagactttatcaataagagcaatgattt 293


RESULT  37
V73620
ID   V73620 standard; DNA; 464 BP.
XX
AC   V73620;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment CH4001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
```



```
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains  -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 464 BP; 177 A; 94 C; 69 G; 124 T; 0 other;


  Query Match             0.8%;  Score 49;  DB 19;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 2e-11;
```



```
  Matches   49;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     764 ccaaaaagaatcagtatttttcagactttatcaataagagcaatgattt 812
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 ccaaaaagaatcagtatttttcagactttatcaataagagcaatgattt 293


RESULT  38
V73619
ID   V73619 standard; DNA; 464 BP.
XX
AC   V73619;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cagA DNA fragment CH2001C2.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains  -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 18; Figure 10; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
```

CC strain according to the VDG alleles present in that particular H. pylori
CC strain. The virulence determinant genes are the genetic elements involved
CC in enabling, determining, and marking the infectivity and/or
CC pathogenicity of the H. pylori strain. The method provides a way of
CC detecting H. pylori strains in a sample with respect to the development
CC of chronic active gastritis, gastric and duodenal ulcers, gastric
CC adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC determining eradication therapy. V73508-V73546 represent PCR primers
CC and probes used in the detection of the H. pylori vacA and cagA genes.
CC The primers and probes are used especially to detect the vacA S regions
CC S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC V73547-V73785.
XX
SQ Sequence 464 BP; 177 A; 96 C; 70 G; 121 T; 0 other;

Query Match 0.7%; Score 40; DB 19; Length 464;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 763 accaaaaagaatcagtatttttcagactttatcaataaga 802
||||||||||||||||||||||||||||||||||||||||
Db 244 accaaaaagaatcagtatttttcagactttatcaataaga 283

RESULT 39
V73622
ID V73622 standard; DNA; 464 BP.
XX
AC V73622;
XX
DT 22-MAR-2000 (first entry)
XX
DE H. pylori cagA DNA fragment 8828001C2.
XX
KW PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS Helicobacter pylori.
XX
PN WO9816658-A2.
XX
PD 23-APR-1998.
XX
PF 10-OCT-1997; 97WO-EP05614.
XX
PR 16-OCT-1996; 96EP-0870131.
PR 09-SEP-1997; 97EP-0870133.

XX
PA (INNO-) INNOGENETICS NV.
PA (DDLD-) DDL BV.
XX
PI Quint W, Van Doorn L;
XX
DR WPI; 1998-251300/22.
XX
PT Method for detecting and/or typing Helicobacter pylori strains -
PT comprises use of primers and probes based on vacA and cagA gene
XX
PS Claim 18; Figure 10; 122pp; English.
XX
CC This invention describes a novel method for the detection and/or typing
CC of Helicobacter pylori strains present in a sample using PCR primers and
CC probes to detect regions of the vacuolating toxin (vacA) gene and other
CC virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC gene. The method allows the typing and allele-specific detection of a
CC strain according to the VDG alleles present in that particular H. pylori
CC strain. The virulence determinant genes are the genetic elements involved
CC in enabling, determining, and marking the infectivity and/or
CC pathogenicity of the H. pylori strain. The method provides a way of
CC detecting H. pylori strains in a sample with respect to the development
CC of chronic active gastritis, gastric and duodenal ulcers, gastric
CC adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC determining eradication therapy. V73508-V73546 represent PCR primers
CC and probes used in the detection of the H. pylori vacA and cagA genes.
CC The primers and probes are used especially to detect the vacA S regions
CC S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC V73547-V73785.
XX
SQ Sequence 464 BP; 176 A; 92 C; 71 G; 125 T; 0 other;

Query Match 0.6%; Score 38; DB 19; Length 464;
Best Local Similarity 100.0%; Pred. No. 9.6e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 946 atccgaaatttttatggaaaaatatcatacaaccccctat 983
|||||||||||||||||||||||||||||||||||||||
Db 427 atccgaaatttttatggaaaaatatcatacaaccccctat 464

RESULT 40
V73615
ID V73615 standard; DNA; 464 BP.
XX
AC V73615;
XX
DT 22-MAR-2000 (first entry)
XX

DE H. pylori cagA DNA fragment J39001C2.
XX
KW PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS Helicobacter pylori.
XX
PN WO9816658-A2.
XX
PD 23-APR-1998.
XX
PF 10-OCT-1997; 97WO-EP05614.
XX
PR 16-OCT-1996; 96EP-0870131.
PR 09-SEP-1997; 97EP-0870133.
XX
PA (INNO-) INNOGENETICS NV.
PA (DDLD-) DDL BV.
XX
PI Quint W, Van Doorn L;
XX
DR WPI; 1998-251300/22.
XX
PT Method for detecting and/or typing Helicobacter pylori strains -
PT comprises use of primers and probes based on vacA and cagA gene
XX
PS Claim 18; Figure 10; 122pp; English.
XX
CC This invention describes a novel method for the detection and/or typing
CC of Helicobacter pylori strains present in a sample using PCR primers and
CC probes to detect regions of the vacuolating toxin (vacA) gene and other
CC virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC gene. The method allows the typing and allele-specific detection of a
CC strain according to the VDG alleles present in that particular H. pylori
CC strain. The virulence determinant genes are the genetic elements involved
CC in enabling, determining, and marking the infectivity and/or
CC pathogenicity of the H. pylori strain. The method provides a way of
CC detecting H. pylori strains in a sample with respect to the development
CC of chronic active gastritis, gastric and duodenal ulcers, gastric
CC adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC determining eradication therapy. V73508-V73546 represent PCR primers
CC and probes used in the detection of the H. pylori vacA and cagA genes.
CC The primers and probes are used especially to detect the vacA S regions
CC S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC V73547-V73785.
XX
SQ Sequence 464 BP; 177 A; 92 C; 73 G; 122 T; 0 other;

Query Match 0.6%; Score 34; DB 19; Length 464;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 gcttttcttaaagttgataacgctgtcgcttcat 649
||||||||||||||||||||||||||||||||||
Db 97 gcttttcttaaagttgataacgctgtcgcttcat 130

RESULT 41
V45503
ID V45503 standard; DNA; 37 BP.
XX
AC V45503;
XX
DT 15-FEB-1999 (first entry)
XX
DE Helicobacter pylori cagA gene PCR primer.
XX
KW Vaccine; antigen; antigen; toxin; diagnosis; gastritis; ulcer;
KW stomach cancer; cagA gene; PCR; primer; ss..
XX
OS Synthetic.
OS Helicobacter pylori.
XX
PN WO9844130-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-KR00073.
XX
PR 31-MAR-1997; 97KR-0011951.
PR 31-MAR-1997; 97KR-0011950.
XX
PA (DAEW-) DAEWOONG PHARM CO LTD.
XX
PI Choi D, Jung H, Kim B, Park M, Shin S, Yu Y;
XX
DR WPI; 1998-568279/48.
XX
PT New chimeric proteins for use against Helicobacter pylori -
PT comprising an antigenic protein of H. pylori and A1 and B subunits
PT of Vibrio cholerae toxin, preferably produced by recombinant
PT techniques
XX
PS Example 2-2; Page 10; 102pp; English.
XX
CC PCR primers (see V45503 and V45504) are designed for the PCR
CC amplification of the Helicobacter pylori cagA gene. The invention
CC relates to recombinant DNA (see V62460-61) comprising a fusion gene

CC prepared by ligating an antigenic determinant coding gene (e.g. the
CC cagA gene) of H. pylori and A2 and B subunit genes of Vibrio
CC cholerae. Also claimed are chimeric proteins (see W80599-600)
CC encoded by such recombinant DNAs, methods for the recombinant
CC production of the chimeric proteins, and use of the chimeric
CC proteins in preventative and therapeutic vaccines for H. pylori-
CC associated diseases such as gastritis, gastric ulcer, duodenal
CC ulcer and gastric cancer.
XX
SQ Sequence 37 BP; 14 A; 12 C; 6 G; 5 T; 0 other;

Query Match 0.5%; Score 31; DB 19; Length 37;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 535 atgactaacgaaaccattgaccaacaaccac 565
||||||||||||||||||||||||||||||
Db 7 atgactaacgaaaccattgaccaacaaccac 37

RESULT 42
X14282/c
ID X14282 standard; DNA; 990 BP.
XX
AC X14282;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1746 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 43..873
FT /*tag= a
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR P-PSDB; W98563.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 1; Page 1104-1105; 2054pp; English.
XX
CC This aequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 990 BP; 337 A; 148 C; 199 G; 306 T; 0 other;

Query Match 0.5%; Score 31; DB 19; Length 990;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5050 ttacaatttgagccattctttagcttgtttt 5080
||||||||||||||||||||||||||||||
Db 31 TTACAATTTGAGCCATTCTTTAGCTTGTTTT 1

RESULT 43
Z50141
ID Z50141 standard; DNA; 36 BP.
XX
AC Z50141;
XX
DT 04-MAY-2000 (first entry)
XX
DE CAG 1 PCR primer for H. pylori GGT gene amplification.
XX
KW PCR primer; gamma-glutamyl transpeptidase; GGT; antibody; vaccine;
KW Helicobacter infection; chronic gastritis; gastroduodenal ulcer;
KW gastric cancer; probe; antibacterial; antiulcer; antiinflammatory;
KW cytostatic; ss.
XX
OS Helicobacter pylori.
XX

PN WO200001825-A1.
XX
PD 13-JAN-2000.
XX
PF 28-JUN-1999; 99WO-EP04459.
XX
PR 01-JUL-1998; 98US-0108213.
XX
PA (INSP) INST PASTEUR.
XX
PI Chevalier C, Thiberge J, Labigne A;
XX
DR WPI; 2000-182116/16.
XX
PT New isolated bacterial enzymes, used to develop products for the
PT diagnosis, prevention and treatment of Helicobacter infections,
PT particularly H. pylori infections -
XX
PS Example 1; Page 32; 96pp; English.
XX
CC The patent discloses the use of a purified polynucleotide coding for
CC gamma-glutamyl transpeptidase (GGT) from Helicobacter pylori. Protective
CC antibodies induced by the corresponding protein can be used for
CC detecting Helicobacter infection, particularly H. pylori. The
CC polynucleotides can be used for immunising animals or humans against
CC Helicobacter. H. pylori infections that can be treated include, chronic
CC gastritis, gastroduodenal ulcers and gastric cancer. Sequences Z50134-44
CC represent primers used for PCR, cloning and sequencing of the H. pylori
CC chromosomal GGT gene.
XX
SQ Sequence 36 BP; 10 A; 10 C; 9 G; 7 T; 0 other;

Query Match 0.5%; Score 27; DB 21; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 333 tagcctttagacgcctgcaacgatcgg 359
|||||||||||||||||||||||||||
Db 10 tagcctttagacgcctgcaacgatcgg 36

RESULT 44
Z50143
ID Z50143 standard; DNA; 35 BP.
XX
AC Z50143;
XX
DT 04-MAY-2000 (first entry)
XX
DE CAG 3 PCR primer for H. pylori GGT gene amplification.

XX
KW PCR primer; gamma-glutamyl transpeptidase; GGT; antibody; vaccine;
KW Helicobacter infection; chronic gastritis; gastroduodenal ulcer;
KW gastric cancer; probe; antibacterial; antiulcer; antiinflammatory;
KW cytostatic; ss.
XX
OS Helicobacter pylori.
XX
PN WO200001825-A1.
XX
PD 13-JAN-2000.
XX
PF 28-JUN-1999; 99WO-EP04459.
XX
PR 01-JUL-1998; 98US-0108213.
XX
PA (INSP) INST PASTEUR.
XX
PI Chevalier C, Thiberge J, Labigne A;
XX
DR WPI; 2000-182116/16.
XX
PT New isolated bacterial enzymes, used to develop products for the
PT diagnosis, prevention and treatment of Helicobacter infections,
PT particularly H. pylori infections -
XX
PS Example 1; Page 32; 96pp; English.
XX
CC The patent discloses the use of a purified polynucleotide coding for
CC gamma-glutamyl transpeptidase (GGT) from Helicobacter pylori. Protective
CC antibodies induced by the corresponding protein can be used for
CC detecting Helicobacter infection, particularly H. pylori. The
CC polynucleotides can be used for immunising animals or humans against
CC Helicobacter. H. pylori infections that can be treated include, chronic
CC gastritis, gastroduodenal ulcers and gastric cancer. Sequences Z50134-44
CC represent primers used for PCR, cloning and sequencing of the H. pylori
CC chromosomal GGT gene.
XX
SQ Sequence 35 BP; 9 A; 7 C; 12 G; 7 T; 0 other;

Query Match 0.4%; Score 26; DB 21; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3403 gggctttcaaggaatcaagaattggc 3428
||||||||||||||||||||||||||
Db 10 gggctttcaaggaatcaagaattggc 35

RESULT 45

```
V65776/c
ID   V65776 standard; DNA; 28 BP.
XX
AC   V65776;
XX
DT   18-JAN-1999  (first entry)
XX
DE   Helicobacter pylori cagA gene primer R009.
XX
KW   Porin; porA gene; cytotoxin; infection; diagnosis; vaccine;
KW   Campylobacter jejuni; cagA gene; PCR; primer; ss.
XX
OS   Synthetic.
OS   Helicobacter pylori.
XX
PN   WO9842842-A1.
XX
PD   01-OCT-1998.
XX
PF   25-MAR-1998;   98WO-CA00272.
XX
PR   25-MAR-1997;   97US-0041200.
XX
PA   (CNDG ) CANADA MIN HEALTH & WELFARE.
XX
PI   Bacon DJ,  Bolla J,  Johnson WM,  Rodgers F;
XX
DR   WPI; 1998-532008/45.
XX
PT   New nucleic acid encoding the porA porin from Campylobacter jejuni -
PT   useful for detecting C. jejuni infections in humans and animals
XX
PS   Example 2; Page 51; 131pp; English.
XX
CC   Primers DZ3 and R009 (see V65775-76) are specific for the cagA
CC   gene of Helicobacter pylori.  These primers, along with other
CC   primers specific for the Escherichia coli VT1 gene (see V65771-72),
CC   the E. coli VT2 gene (see V65773-74), or the H. pylori vacA gene
CC   (see V65777-78), were used in the PCR amplification of
CC   Campylobacter jejuni strain 2483 genomic DNA.  No amplicons were
CC   generated.  The C. jejuni strain 2483 porin (see W79959) shows
CC   homology to another isolated C. jejuni cytotoxic porin protein,
CC   to the C. jejuni major outer membrane protein, and to the 45 and 51
CC   kDa outer membrane proteins of Wolinella recta.  The isolated
CC   porA gene (see V65769) of strain 2483, and porin polypeptides, can
CC   be used in methods for detecting C. jejuni infections in humans
CC   and animals.  Immunogenically active peptides of porin are used in
CC   vaccines.
XX
SQ   Sequence 28 BP; 8 A; 5 C; 5 G; 10 T; 0 other;
```



```
  Query Match             0.4%;  Score 25;  DB 19;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 0.4;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1816 ttccaaaaagactctaaggcttatt 1840
          |||||||||||||||||||||||||
Db     25 TTCCAAAAAGACTCTAAGGCTTATT 1

RESULT  46
Z50142/c
ID   Z50142 standard; DNA; 34 BP.
XX
AC   Z50142;
XX
DT   04-MAY-2000  (first entry)
XX
DE   CAG 2 PCR primer for H. pylori GGT gene amplification.
XX
KW   PCR primer; gamma-glutamyl transpeptidase; GGT; antibody; vaccine;
KW   Helicobacter infection; chronic gastritis; gastroduodenal ulcer;
KW   gastric cancer; probe; antibacterial; antiulcer; antiinflammatory;
KW   cytostatic; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO200001825-A1.
XX
PD   13-JAN-2000.
XX
PF   28-JUN-1999;   99WO-EP04459.
XX
PR   01-JUL-1998;   98US-0108213.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Chevalier C,  Thiberge J,  Labigne A;
XX
DR   WPI; 2000-182116/16.
XX
PT   New isolated bacterial enzymes, used to develop products for the
PT   diagnosis, prevention and treatment of Helicobacter infections,
PT   particularly H. pylori infections -
XX
PS   Example 1; Page 32; 96pp; English.
XX
CC   The patent discloses the use of a purified polynucleotide coding for
CC   gamma-glutamyl transpeptidase (GGT) from Helicobacter pylori. Protective
CC   antibodies induced by the corresponding protein can be used for
CC   detecting Helicobacter infection, particularly H. pylori. The
```



```
CC   polynucleotides can be used for immunising animals or humans against
CC   Helicobacter. H. pylori infections that can be treated include, chronic
CC   gastritis, gastroduodenal ulcers and gastric cancer. Sequences Z50134-44
CC   represent primers used for PCR, cloning and sequencing of the H. pylori
CC   chromosomal GGT gene.
XX
SQ   Sequence 34 BP; 10 A; 8 C; 7 G; 9 T; 0 other;

  Query Match             0.4%;  Score 25;  DB 21;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 0.39;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    823 gacaatctcattgatgtagaatctt 847
          |||||||||||||||||||||||||
Db     34 GACAATCTCATTGATGTAGAATCTT 10

RESULT  47
V73519
ID   V73519 standard; DNA; 24 BP.
XX
AC   V73519;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cag A primer cagF.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
```



```
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 2; Page 44; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas; and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 24 BP; 11 A; 8 C; 3 G; 2 T; 0 other;

  Query Match             0.4%;  Score 24;  DB 19;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    551 ttgaccaacaaccacaaaccgaag 574
          ||||||||||||||||||||||||
Db      1 ttgaccaacaaccacaaaccgaag 24

RESULT  48
V45504/c
ID   V45504 standard; DNA; 30 BP.
XX
AC   V45504;
XX
DT   15-FEB-1999  (first entry)
XX
DE   Helicobacter pylori cagA gene PCR primer.
XX
KW   Vaccine; antigen; antigen; toxin; diagnosis; gastritis; ulcer;
KW   stomach cancer; cagA gene; PCR; primer; ss.
XX
OS   Synthetic.
OS   Helicobacter pylori.
```

XX
PN WO9844130-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-KR000073.
XX
PR 31-MAR-1997; 97KR-0011951.
PR 31-MAR-1997; 97KR-0011950.
XX
PA (DAEW-) DAEWOONG PHARM CO LTD.
XX
PI Choi D, Jung H, Kim B, Park M, Shin S, Yu Y;
XX
DR WPI; 1998-568279/48.
XX
PT New chimeric proteins for use against Helicobacter pylori -
PT comprising an antigenic protein of H. pylori and A1 and B subunits
PT of Vibrio cholerae toxin, preferably produced by recombinant
PT techniques
XX
PS Example 2-2; Page 10; 102pp; English.
XX
CC PCR primers (see V45503 and V45504) are designed for the PCR
CC amplification of the Helicobacter pylori cagA gene. The invention
CC relates to recombinant DNA (see V62460-61) comprising a fusion gene
CC prepared by ligating an antigenic determinant coding gene (e.g. the
CC cagA gene) of H. pylori and A2 and B subunit genes of Vibrio
CC cholerae. Also claimed are chimeric proteins (see W80599-600)
CC encoded by such recombinant DNAs, methods for the recombinant
CC production of the chimeric proteins, and use of the chimeric
CC proteins in preventative and therapeutic vaccines for H. pylori-
CC associated diseases such as gastritis, gastric ulcer, duodenal
CC ulcer and gastric cancer.
XX
SQ Sequence 30 BP; 10 A; 5 C; 4 G; 11 T; 0 other;

Query Match 0.4%; Score 23; DB 19; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3956 aaggtggtttccaaaaatcttaa 3978
||||||||||||||||||||||
Db 30 AAGGTGGTTTCCAAAAATCTTAA 8

RESULT 49
X14522
ID X14522 standard; DNA; 2498 BP.
XX

AC X14522;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1302 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 91..2448
FT /*tag= a
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR P-PSDB; W98803.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 1; Page 1757-1761; 2054pp; English.
XX
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 2498 BP; 806 A; 455 C; 596 G; 641 T; 0 other;

Query Match 0.4%; Score 23; DB 19; Length 2498;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4512 caaaaaatcatagagcttttaga 4534
|||||||||||||||||||||||
Db 190 caaaaaatcatagagcttttaga 212

RESULT 50
V73520/c
ID V73520 standard; DNA; 22 BP.
XX
AC V73520;
XX
DT 22-MAR-2000 (first entry)
XX
DE H. pylori cag A primer cagR.
XX
KW PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW S region; S1a; S1b; S1c; S2; M region; M1; M2; as.
XX
OS Helicobacter pylori.
XX
PN WO9816658-A2.
XX
PD 23-APR-1998.
XX
PF 10-OCT-1997; 97WO-EP05614.
XX
PR 16-OCT-1996; 96EP-0870131.
PR 09-SEP-1997; 97EP-0870133.
XX
PA (INNO-) INNOGENETICS NV.
PA (DDLD-) DDL BV.
XX
PI Quint W, Van Doorn L;
XX
DR WPI; 1998-251300/22.
XX
PT Method for detecting and/or typing Helicobacter pylori strains -
PT comprises use of primers and probes based on vacA and cagA gene
XX
PS Claim 2; Page 44; 122pp; English.
XX
CC This invention describes a novel method for the detection and/or typing
CC of Helicobacter pylori strains present in a sample using PCR primers and
CC probes to detect regions of the vacuolating toxin (vacA) gene and other

CC virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC gene. The method allows the typing and allele-specific detection of a
CC strain according to the VDG alleles present in that particular H. pylori
CC strain. The virulence determinant genes are the genetic elements involved
CC in enabling, determining, and marking the infectivity and/or
CC pathogenicity of the H. pylori strain. The method provides a way of
CC detecting H. pylori strains in a sample with respect to the development
CC of chronic active gastritis, gastric and duodenal ulcers, gastric
CC adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC determining eradication therapy. V73508-V73546 represent PCR primers
CC and probes used in the detection of the H. pylori vacA and cagA genes.
CC The primers and probes are used especially to detect the vacA S regions
CC S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC V73547-V73785.
XX
SQ Sequence 22 BP; 4 A; 7 C; 3 G; 8 T; 0 other;

Query Match 0.4%; Score 22; DB 19; Length 22;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 712 ggaatctcgcaattaagggaag 733
||||||||||||||||||||||
Db 22 GGAATCTCGCAATTAAGGGAAG 1

RESULT 51
V90656
ID V90656 standard; DNA; 267 BP.
XX
AC V90656;
XX
DT 18-FEB-1999 (first entry)
XX
DE Nucleotide sequence of clone Y178-1T7.SEQ from cluster 18.
XX
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX
OS Helicobacter pylori.
XX
PN WO9849314-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-US08487.
XX
PR 14-OCT-1997; 97US-0061958.
PR 25-APR-1997; 97US-0045107.
XX

```
PA   (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI   Chow TP,  Fry KE,  Lim MY,  McAtee CP;
XX
DR   WPI; 1999-009433/01.
XX
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
XX
PS   Claim 27; Page 144; 402pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori antigenic protein
CC   that is characterised by immunoreactivity with H. pylori-positive
CC   antisera. The proteins are highly immunogenic and induce a long-lasting
CC   immune response that persists even after antimicriobial treatment. In
CC   antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC   highly sensitive and specific. The specification also describes 69
CC   previously unrecognised immunogenic cluster families. H. pylori antigens
CC   are used to detect H. pylori-specific antibodies, for diagnosing
CC   infection or to confirm eradication of infection, and in vaccines to
CC   protect against H. pylori infection and related diseases (gastritis,
CC   peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ   Sequence 267 BP; 101 A; 39 C; 33 G; 94 T; 0 other;


  Query Match             0.4%;  Score 22;  DB 20;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     427 aatgttcaaagacatgaattga 448
           ||||||||||||||||||||||
Db     238 aatgttcaaagacatgaattga 259


RESULT  52
V73527/c
ID   V73527 standard; DNA; 21 BP.
XX
AC   V73527;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cag A primer cagSR.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
```



```
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 2; Page 44; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 21 BP; 4 A; 2 C; 6 G; 9 T; 0 other;


  Query Match             0.4%;  Score 21;  DB 19;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     652 gatcctgatcaaaaaccaatc 672
           |||||||||||||||||||||
```



```
Db      21 GATCCTGATCAAAAACCAATC 1


RESULT  53
V32400
ID   V32400 standard; cDNA to mRNA; 1603 BP.
XX
AC   V32400;
XX
DT   11-SEP-1998  (first entry)
XX
DE   Homo sapiens spry1 gene.
XX
KW   spry1; sprouty 1 protein; angiogenesis; treatment; cancer;
KW   lateral inhibitory signal; epithelial tube branching; inhibition;
KW   neovascularisation; diabetic retinopathy; retrolental fibroplasia;
KW   neovascular glaucoma; psoriasis; angiofibromas; immune; non-immune;
KW   inflammation; rheumatoid arthritis; capillary proliferation;
KW   atherosclerotic plaques; haemangiomas, Kaposi's Sarcoma;
KW   endometriosis; scar formation; wound healing; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..420
FT                   /*tag= a
FT                   /product= sprouty 1 protein
XX
PN   WO9820032-A1.
XX
PD   14-MAY-1998.
XX
PF   07-NOV-1997;   97WO-US20398.
XX
PR   07-NOV-1996;   96US-0030232.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Hacohen N,  Krasnow MA;
XX
DR   WPI; 1998-286862/25.
DR   P-PSDB; W48793.
XX
PT   New isolated human sprouty polypeptide(s) - used to develop products
PT   for treating disorders involving angiogenic dysfunction, e.g.
PT   cancers, diabetic retinopathy, glaucoma, psoriasis or inflammation
XX
PS   Claim 14; Page 50-51; 90pp; English.
XX
CC   The sequence is that encoding human sprouty 1 protein. The protein
CC   functions as a lateral inhibitory signal during epithelial tube
```



```
CC   branching. It can be used for inhibiting angiogenesis and
CC   neovascularisation in the treatment of e.g. cancers, diabetic
CC   retinopathy, retrolental fibroplasia, neovascular glaucoma, psoriasis,
CC   angiofibromas, immune and non-immune inflammation (including rheumatoid
CC   arthritis), capillary proliferation within atherosclerotic plaques,
CC   haemangiomas, Kaposi's Sarcoma, endometriosis and unwanted scar
CC   formation in wound healing.  Products of the DNA and proteins
CC   can also be used for detection and diagnosis.
XX
SQ   Sequence 1603 BP; 411 A; 320 C; 316 G; 556 T; 0 other;


  Query Match             0.4%;  Score 21;  DB 19;  Length 1603;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db    1579 attaatgaaaaaaaaaaaaa 1599


RESULT  54
A14074
ID   A14074 standard; cDNA; 2030 BP.
XX
AC   A14074;
XX
DT   21-JUL-2000  (first entry)
XX
DE   Human SPROUTY-1 partial cDNA, SEQ ID NO:8.
XX
KW   SPROUTY-1; spry1; fibroblast growth factor; FGF inhibitor;
KW   signalling; tumorigenesis; neurogenesis; metanephephrogenesis; cancer;
KW   tumour; prostate; adrenal gland; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..889
FT                   /*tag= a
FT                   /partial
FT                   /product= "Human SPROUTY-1 protein fragment, SEQ ID NO:7
FT                            (Y87594)"
FT                   /note= "No initiation codon given in the specification"
FT   Misc_feature    574..576
FT                   /note= "The bases encoding residues 191-258 of the
FT                         corresponding protein are missing from this
FT                         sequence"
XX
PN   WO200015781-A1.
XX
```

```
PD   23-MAR-2000.
XX
PF   10-SEP-1999;   99WO-US20638.
XX
PR   11-SEP-1998;   98US-0099936.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Edmonds BT;
XX
DR   WPI; 2000-271422/23.
DR   P-PSDB; Y87594.
XX
PT   New polynucleotide encoding a 319 amino acid sequence, used as a
PT   fibroblast growth factor (FGF) antagonist for treating prostate cancer
PT   and inhibiting angiogenesis in mammals -
XX
PS   Example 1; Fig 8; 93pp; English.
XX
CC   This sequence represents a cDNA clone encoding a human SPROUTY-1 protein
CC   (also referred to as spryl in the specification). SPROUTY-1 proteins
CC   (Y87591-Y87595) are antagonists of members of the fibroblast growth
CC   factor (FGF) family. The SPROUTY-1 proteins of the invention, and
CC   fragments thereof may be used in mammals to limit FGF levels in vivo and
CC   to control FGF-mediated signalling. They may be used to counteract FGF
CC   overexpression associated with certain types of tumours, and to control
CC   the growth, development or differentiation of any cell responsive to
CC   FGF-mediated signalling. For example, SPROUTY-1 proteins may be used to
CC   control tumorigenesis, neurogenesis, metanephrogenesis, and particularly
CC   angiogenesis. They may also be used to treat conditions associated with
CC   SPROUTY-1 underexpression including polycystic kidney disease,
CC   nephropathy, lung fibrosis, hepatic fibrogenesis, endometrial
CC   insufficiency and various cancers, particularly prostatic tumours.
CC   They may additionally be used to treat a mammal with an adrenal gland
CC   pathology involving hypersensitive responses to FGF. Nucleic acid
CC   sequences encoding SPROUTY-1 proteins may be used to recombinantly
CC   express SPROUTY-1 proteins, as a source of primers and probes, and for
CC   antisense suppression of SPROUTY-1 gene expression.
XX
SQ   Sequence 2030 BP; 543 A; 435 C; 407 G; 645 T; 0 other;

  Query Match             0.4%;  Score 21;  DB 21;  Length 2030;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db    2010 attaatgaaaaaaaaaaaaa 2030
```



```
RESULT  55
A14073
ID   A14073 standard; cDNA; 2138 BP.
XX
AC   A14073;
XX
DT   21-JUL-2000  (first entry)
XX
DE   Human SPROUTY-1 partial cDNA, SEQ ID NO:6.
XX
KW   SPROUTY-1; spryl; fibroblast growth factor; FGF inhibitor;
KW   signalling; tumorigenesis; neurogenesis; metanephephrogenesis; cancer;
KW   tumour; prostate; adrenal gland; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..889
FT                   /*tag= a
FT                   /partial
FT                   /product= "Human SPROUTY-1 protein fragment, SEQ ID NO:5
FT                    (Y87593)"
FT                   /note= "No initiation codon given in the specification"
XX
PN   WO200015781-A1.
XX
PD   23-MAR-2000.
XX
PF   10-SEP-1999;   99WO-US20638.
XX
PR   11-SEP-1998;   98US-0099936.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Edmonds BT;
XX
DR   WPI; 2000-271422/23.
DR   P-PSDB; Y87593.
XX
PT   New polynucleotide encoding a 319 amino acid sequence, used as a
PT   fibroblast growth factor (FGF) antagonist for treating prostate cancer
PT   and inhibiting angiogenesis in mammals -
XX
PS   Example 1; Fig 6; 93pp; English.
XX
CC   This sequence represents a cDNA clone encoding a human SPROUTY-1 protein
CC   (also referred to as spryl in the specification). SPROUTY-1 proteins
CC   (Y87591-Y87595) are antagonists of members of the fibroblast growth
CC   factor (FGF) family. The SPROUTY-1 proteins of the invention, and
CC   fragments thereof may be used in mammals to limit FGF levels in vivo and
CC   to control FGF-mediated signalling. They may be used to counteract FGF
```



```
CC   overexpression associated with certain types of tumours, and to control
CC   the growth, development or differentiation of any cell responsive to
CC   FGF-mediated signalling. For example, SPROUTY-1 proteins may be used to
CC   control tumorigenesis, neurogenesis, metanephrogenesis, and particularly
CC   angiogenesis. They may also be used to treat conditions associated with
CC   SPROUTY-1 underexpression including polycystic kidney disease,
CC   nephropathy, lung fibrosis, hepatic fibrogenesis, endometrial
CC   insufficiency and various cancers, particularly prostatic tumours.
CC   They may additionally be used to treat a mammal with an adrenal gland
CC   pathology involving hypersensitive responses to FGF. Nucleic acid
CC   sequences encoding SPROUTY-1 proteins may be used to recombinantly
CC   express SPROUTY-1 proteins, as a source of primers and probes, and for
CC   antisense suppression of SPROUTY-1 gene expression.
XX
SQ   Sequence 2138 BP; 590 A; 454 C; 433 G; 661 T; 0 other;

  Query Match             0.4%;  Score 21;  DB 21;  Length 2138;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db    2084 attaatgaaaaaaaaaaaaa 2104


RESULT  56
V25297
ID   V25297 standard; DNA; 20 BP.
XX
AC   V25297;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Primer F2 for H.pylori MurI gene.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   PCR primer; ss.
XX
OS   Synthetic.
OS   Helicobacter pylori.
XX
PN   WO9737044-A1.
XX
PD   09-OCT-1997.
XX
PF   27-MAR-1997;   97WO-US05223.
XX
PR   06-DEC-1996;   96US-0761318.
```



```
PR   29-MAR-1996;   96US-0625811.
PR   02-APR-1996;   96US-0758731.
PR   25-OCT-1996;   96US-0736905.
PR   28-OCT-1996;   96US-0738859.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Alm RA,  Smith D;
XX
DR   WPI; 1997-503122/46.
XX
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
XX
PS   Example; Page 109; 1145pp; English.
XX
CC   This sequence represents a primer for the H.pylori MurI gene.
CC   The amplified sequence was used to compare homology of the coding
CC   sequences of the invention with other known proteins. The protein encoded
CC   by the DNA of the invention may be used in a vaccine to prevent or treat
CC   H.pylori infection or to identify H.pylori polypeptide binding compounds,
CC   useful as potential H.pylori life cycle activators or inhibitors. The DNA
CC   and probes derived from it may be used for the identification of H.pylori
CC   in a sample and the diagnosis of H.pylori infection. Nucleic acid
CC   sequences complementary to the DNA act as antisense sequences and can be
CC   used to prevent the translation of H.pylori mRNA. Antibodies against the
CC   protein can be used in immunoassays to evaluate the abundance and
CC   distribution of H.pylori-specific antigens. The genomic sequence of
CC   H.pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H.pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H.pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
XX
SQ   Sequence 20 BP; 5 A; 4 C; 8 G; 3 T; 0 other;

  Query Match             0.3%;  Score 20;  DB 18;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5184 cgcccgaatggatgagtagg 5203
           ||||||||||||||||||||
Db       1 cgcccgaatggatgagtagg 20
```

```
RESULT  57
V73508
ID   V73508 standard; DNA; 20 BP.
XX
AC   V73508;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cag A probe cagApro DNA.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 3; Page 45; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
```



```
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 20 BP; 3 A; 5 C; 5 G; 7 T; 0 other;


  Query Match             0.3%;  Score 20;  DB 19;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    628 gttgataacgctgtcgcttc 647
          ||||||||||||||||||||
Db      1 gttgataacgctgtcgcttc 20


RESULT  58
V73534/c
ID   V73534 standard; DNA; 20 BP.
XX
AC   V73534;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cag A probe cagprobe3 DNA.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816658-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
```



```
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 3; Page 46; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC   V73547-V73785.
XX
SQ   Sequence 20 BP; 3 A; 2 C; 4 G; 11 T; 0 other;


  Query Match             0.3%;  Score 20;  DB 19;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    742 aataaagcgatcaaaaatcc 761
          ||||||||||||||||||||
Db     20 AATAAAGCGATCAAAAATCC 1


RESULT  59
V50362/c
ID   V50362 standard; DNA; 20 BP.
XX
AC   V50362;
XX
DT   20-OCT-1998  (first entry)
XX
DE   Antisense oligomer for inhibiting Helicobacter pylori infection #6.
XX
KW   Antisense oligomer; Helicobacter pylori; H. pylori; inhibition;
KW   infection; morpholino backbone; ss.
XX
OS   Synthetic.
```



```
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   modified_site   1..20
FT                   /*tag=  a
FT                   /note= "morpholino backbone"
XX
PN   WO9832467-A2.
XX
PD   30-JUL-1998.
XX
PF   23-JAN-1998;   98WO-US01393.
XX
PR   30-APR-1997;   97US-0848844.
PR   24-JAN-1997;   97US-0036366.
XX
PA   (ANTI-) ANTIVIRALS INC.
PA   (UYNE-) UNIV NEBRASKA.
XX
PI   Bragg R,  Iverson PL,  Summerton JE,  Weller DD;
XX
DR   WPI; 1998-427687/36.
XX
PT   Antisense conjugates useful in treating H. pylori infection -
PT   include an antisense oligomer and transport moiety which facilitates
PT   uptake by active transport or pH gradient transport
XX
PS   Claim 4; Page 7; 36pp; English.
XX
CC   A method has been developed for treating Helicobacter pylori infection.
CC   The method comprises administering to the subject, by oral route, a
CC   therapeutically effective amount of an antisense conjugate composed of:
CC   (a) a nuclease-resistant antisense oligomer, which can inhibit H. pylori
CC   infection by base-specific Watson-Crick binding to a H. pylori mRNA
CC   transcript, and (b) a transport moiety (TM) conjugated to the antisense
CC   moiety. The TM facilitates uptake of the composition from the environment
CC   of the stomach into the cytoplasm of H. pylori cells by active transport
CC   or pH-gradient transport across the cell membrane. V50357 to V50367
CC   represent specifically claimed antisense oligomers for use in the method
CC   of the invention. The conjugate may be used in the treatment of H. pylori
CC   infection, including chronic superficial gastritis, peptic ulcer disease
CC   and chronic atrophic gastritis leading to gastric adenocarcinoma.
CC   Administration is oral, especially in a swellable polymer bolus designed
CC   for sustained release of the conjugate.  The total daily dose is 5-100
CC   mu moles of conjugate. The conjugate is typically administered over
CC   15-20 days, until eradication or a marked reduction of infection is
CC   observed.
XX
SQ   Sequence 20 BP; 2 A; 4 C; 3 G; 11 T; 0 other;
```

```
  Query Match          0.3%;  Score 20;  DB 19;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 54;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    524 aaggagaaacaatgactaac 543
          ||||||||||||||||||||
Db     20 AAGGAGAAACAATGACTAAC 1


RESULT  60
V25299
ID   V25299 standard; DNA; 21 BP.
XX
AC   V25299;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Primer F3 for H.pylori MurI gene.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   PCR primer; ss.
XX
OS   Synthetic.
OS   Helicobacter pylori.
XX
PN   WO9737044-A1.
XX
PD   09-OCT-1997.
XX
PF   27-MAR-1997;   97WO-US05223.
XX
PR   06-DEC-1996;   96US-0761318.
PR   29-MAR-1996;   96US-0625811.
PR   02-APR-1996;   96US-0758731.
PR   25-OCT-1996;   96US-0736905.
PR   28-OCT-1996;   96US-0738859.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Alm RA,  Smith D;
XX
DR   WPI; 1997-503122/46.
XX
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
XX
PS   Example; Page 109; 1145pp; English.
XX
```



```
CC   This sequence represents a primer for the H.pylori MurI gene.
CC   The amplified sequence was used to compare homology of the coding
CC   sequences of the invention with other known proteins. The protein encoded
CC   by the DNA of the invention may be used in a vaccine to prevent or treat
CC   H.pylori infection or to identify H.pylori polypeptide binding compounds,
CC   useful as potential H.pylori life cycle activators or inhibitors. The DNA
CC   and probes derived from it may be used for the identification of H.pylori
CC   in a sample and the diagnosis of H.pylori infection. Nucleic acid
CC   sequences complementary to the DNA act as antisense sequences and can be
CC   used to prevent the translation of H.pylori mRNA. Antibodies against the
CC   protein can be used in immunoassays to evaluate the abundance and
CC   distribution of H.pylori-specific antigens. The genomic sequence of
CC   H.pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H.pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H.pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
XX
SQ   Sequence 21 BP; 3 A; 1 C; 5 G; 12 T; 0 other;

  Query Match          0.3%;  Score 20;  DB 18;  Length 21;
  Best Local Similarity  100.0%;  Pred. No. 54;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5141 tttttaagggcgtatttttg 5160
          ||||||||||||||||||||
Db      1 tttttaagggcgtatttttg 20


RESULT  61
Q38651
ID   Q38651 standard; DNA; 284 BP.
XX
AC   Q38651;
XX
DT   07-JUL-1993  (first entry)
XX
DE   Truncated Rs-AFP2 cDNA.
XX
KW   Raphanus sativus; Brassica; Arabidopsis; Cnicus; Lathyrus; Clitoria;
KW   fungicide; bacteriocide; antibiotic; antifungal; gram positive;
KW   plant disease resistance; low toxicity.
XX
OS   Raphanus sativus.
XX
```



```
FH   Key             Location/Qualifiers
FT   CDS             1..108
FT                   /*tag=  a
XX
PN   WO9305153-A.
XX
PD   18-MAR-1993.
XX
PF   27-AUG-1992;   92WO-GB01570.
XX
PR   29-AUG-1991;   91GB-0018523.
PR   13-FEB-1992;   92GB-0003038.
PR   25-JUN-1992;   92GB-0013526.
XX
PA   (ICIL ) IMPERIAL CHEM IND PLC.
XX
PI   Broekaert WF,  Cammue BPA,  Osborn RW,  Rees SB,  Terras FRG;
PI   Vanderleyden J;
XX
DR   WPI; 1993-100978/12.
XX
PT   Biocidal proteins isolated from seeds of plants - e.g. brassica
PT   or dahlia, useful for increasing plants' resistance to fungal and
PT   bacterial diseases
XX
PS   Example 21; Fig 35; 110pp; English.
XX
CC   This cDNA represents the sequence of Rs-AFP1 from Raphanus sativus.
CC   PCR primer Q38640 was used together with Q38641 to generate a probe
CC   for screening a Raphanus sativus seed cDNA library. This primer
CC   corresponds to amino acids 2 to 7 of Rs-AFP1 and has a sense
CC   orientation. The 144bp product was partially re-amplified using
CC   Q38642 and Q38641 to give a  123bp  product, which was further
CC   reamplified with the same primers and digoxigenin-11-dUTP instead of
CC   dTTP to give a digoxigenin labeled PCR product. This was used to
CC   screen a lambda ZAPII cDNA library by in situ plaque hybridisation.
CC   Positive plaques were purified and subjected to two additional
CC   screening rounds with the same probe. Inserts were excised in vivo
CC   into the pBluescript phagemid form with the aid of helper phage
CC   R408.Inserts from 22 positive clones were excised by EcoRI
CC   digestion and their size compared by agarose gel electrophoresis.
CC   Four clones had insert sizes of approx. 400bp the others between
CC   250-300bp. The inserts of the 4 largest clones were then sequenced
CC   and found to differ only in the length of their 5' and 3' UTR's.
CC   The longest sequence was identified as Rs-AFP1 (Q38650) and the
CC   smallest as Rs-AFP2
XX
SQ   Sequence 284 BP; 89 A; 46 C; 56 G; 93 T; 0 other;

  Query Match          0.3%;  Score 20;  DB 14;  Length 284;
```



```
  Best Local Similarity  100.0%;  Pred. No. 45;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5316 gaaataataacgcatgcaag 5335
          ||||||||||||||||||||
Db      2 gaaataataacgcatgcaag 21


RESULT  62
Q70129
ID   Q70129 standard; cDNA; 284 BP.
XX
AC   Q70129;
XX
DT   14-FEB-1995  (first entry)
XX
DE   Antimicrobial Rs-AFP2.
XX
KW   Antimicrobial; Rs-AFP2; symbiosis; disease-resistance;
KW   fungus-resistance; Clavibacter xyli subsp. cynodontis; Cxc;
KW   crop improvement; endophyte; ss.
XX
OS   Raphanus sativus.
XX
PN   WO9416076-A.
XX
PD   21-JUL-1994.
XX
PF   05-JAN-1994;   94WO-GB00012.
XX
PR   08-JAN-1993;   93GB-0000281.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Dubock AC,  Powell KA,  Rees SB;
XX
DR   WPI; 1994-249223/30.
DR   P-PSDB; R57326.
XX
PT   Antimicrobial protein producing endo-symbiotic microorganisms -
PT   is produced by combining nucleic acids encoding the protein with
PT   an endophyte, useful for protecting plant hosts from esp. fungal
PT   disease
XX
PS   Disclosure; Page 32; 39pp; English.
XX
CC   Plant-derived antimicrobial proteins are expressed in endosymbiotic
CC   Clavibacter xyli subsp. cynodontis (Cxc).  Plants or seeds treated
CC   with recombinant Cxc are protected against fungal disease.  A
CC   suitable antimicrobial protein is Rs-AFP1 from R. sativus.  The
CC   full-length cDNA sequence of Rs-AFP2 is given in Q70129.
```

XX
SQ Sequence 284 BP; 89 A; 46 C; 56 G; 93 T; 0 other;

Query Match 0.3%; Score 20; DB 15; Length 284;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5316 gaaataataacgcatgcaag 5335
||||||||||||||||||||
Db 2 gaaataataacgcatgcaag 21

RESULT 63
A06488/c
ID A06488 standard; cDNA; 302 BP.
XX
AC A06488;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:255.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX

PS Claim 1; Page 183; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which expreas
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. A06241 to A06691 and
CC Y82000 to Y82020 represent sequences used in the exemplification of the
CC present invention.
XX
SQ Sequence 302 BP; 112 A; 38 C; 51 G; 101 T; 0 other;

Query Match 0.3%; Score 20; DB 21; Length 302;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 38 TTAATGAAAAAAAAAAAAAA 19

RESULT 64
X61423
ID X61423 standard; cDNA; 610 BP.
XX
AC X61423;
XX
DT 14-JUL-1999 (first entry)
XX
DE DNA encoding a human secreted protein.
XX
KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease;
KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
KW lung disease; thymus disease; digestive disorder; endocrine disorder;
KW infection; AIDS; ss.
XX
OS Homo sapiens.
XX
PN WO9922243-A1.
XX

PD 06-MAY-1999.
XX
PF 23-OCT-1998; 98WO-US22376.
XX
PR 24-OCT-1997; 97US-0063387.
PR 24-OCT-1997; 97US-0062784.
PR 24-OCT-1997; 97US-0063088.
PR 24-OCT-1997; 97US-0063089.
PR 24-OCT-1997; 97US-0063090.
PR 24-OCT-1997; 97US-0063091.
PR 24-OCT-1997; 97US-0063092.
PR 24-OCT-1997; 97US-0063097.
PR 24-OCT-1997; 97US-0063098.
PR 24-OCT-1997; 97US-0063099.
PR 24-OCT-1997; 97US-0063100.
PR 24-OCT-1997; 97US-0063101.
PR 24-OCT-1997; 97US-0063109.
PR 24-OCT-1997; 97US-0063110.
PR 24-OCT-1997; 97US-0063111.
PR 24-OCT-1997; 97US-0063148.
PR 24-OCT-1997; 97US-0063386.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
PI Feng P, Florence C, Florence KA, Greene JM, Janat F;
PI Kayw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX
DR WPI; 1999-303069/25.
DR P-PSDB; Y19543.
XX
PT New isolated human genes and the secreted polypeptides they encode
XX
PS Claim 3; Page 363; 546pp; English.
XX
CC The specification describes cDNA sequences (X61322-X61470) encoding human
CC secreted proteins (Y19442-Y19590). The polynucleotides and their
CC corresponding secreted polypeptides are useful for preventing, treating
CC or ameliorating medical conditions, e.g. by protein or gene therapy.
CC Pathological conditions can also be diagnosed by determining the amount
CC of the polypeptides in a sample or by determining the presence of
CC mutations in the polynucleotides. Specific uses are described for each
CC of the polynucleotides, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and fetal deficiencies, blood disorders, leukemias,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
CC disorders involving osteoclasts such as osteoporosis, arthritis or

CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
CC disorders, infections and AIDS. The polypeptides are also useful for
CC identifying their binding partners.
XX
SQ Sequence 610 BP; 187 A; 121 C; 103 G; 196 T; 3 other;

Query Match 0.3%; Score 20; DB 20; Length 610;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 566 ttaatgaaaaaaaaaaaaaa 585

RESULT 65
A26393
ID A26393 standard; cDNA; 619 BP.
XX
AC A26393;
XX
DT 29-JUN-2000 (first entry)
XX
DE Human secreted protein gene 48 SEQ ID NO:58.
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour; ss.
XX
OS Homo sapiens.
XX
PN WO200006698-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-US17130.
XX
PR 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;

```
XX
DR   WPI; 2000-195282/17.
DR   P-PSDB; Y91498.
XX
PT   New isolated human genes and the secreted polypeptides they encode,
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 405-406; 634pp; English.
XX
CC   The polynucleotide sequences given in A26346 to A26458 encode the human
CC   secreted proteins given in Y91451 to Y91691. The human secreted proteins
CC   can have activities based on the tissues and cells they are expressed in.
CC   Examples of the activities are: cytostatic; immunosuppressive; antiHIV;
CC   antiinflammatory; nootropic; neuroprotective; antiallergic; osteopathic;
CC   antiarthritic; antibacterial; antidiabetic; antiasthma; antipsoriatic;
CC   and cardiant. The polynucleotides and their corresponding secreted
CC   proteins are useful for preventing, treating or ameliorating medical
CC   conditions, e.g. by protein or gene therapy. Also pathological conditions
CC   can be diagnosed by determining the amount of the proteins in a sample or
CC   by determining the presence of mutations in the polynucleotides. Specific
CC   uses are described for each of the polynucleotides, based on which
CC   tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, tumours,
CC   neurodegenerative disorders, developmental abnormalities and foetal
CC   deficiencies, blood disorders, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, inflammation,
CC   allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC   osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC   transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC   cardiovascular disorders, reproductive disorders, gastrointestinal
CC   disorders, respiratory disorders and metabolic disorders. The proteins
CC   or polynucleotides can also be used as food additives or preservatives.
CC   The proteins are also useful for identifying their binding partners.
CC   A26337 to A26345 and Y91450 are sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 619 BP; 162 A; 168 C; 110 G; 177 T; 2 other;


  Query Match             0.3%;  Score 20;  DB 21;  Length 619;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db    564 ttaatgaaaaaaaaaaaaaa 583


RESULT  66
A09041
```



```
ID   A09041 standard; cDNA; 1118 BP.
XX
AC   A09041;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human TANGO 195 partial cDNA clone T195Athpb93f1.
XX
KW   TANGO 195; transmembrane protein; modulator; proliferation;
KW   differentiation; cell survival; interleukin-1 receptor superfamily;
KW   gene therapy; anti-inflammatory; anti-allergic; cytostatic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Mat_peptide     159..1118
FT                   /*tag= a
XX
PN   WO200018800-A1.
XX
PD   06-APR-2000.
XX
PF   30-SEP-1999;   99WO-US22818.
XX
PR   30-SEP-1998;   98US-0163523.
XX
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI   Busfield SJ;
XX
DR   WPI; 2000-293110/25.
DR   P-PSDB; Y92182.
XX
PT   New cDNA molecules encoding human TANGO 191 and TANGO 195 proteins,
PT   useful for modulating cellular proliferation, differentiation or cell
PT   survival
XX
PS   Disclosure; Fig 8; 122pp; English.
XX
CC   Human TANGO 191 and 195 are transmembrane proteins, useful as modulating
CC   agents in a regulating a variety of cellular processes, e.g. cellular
CC   proliferation, differentiation and survival. hTANGO 191 cDNA was
CC   isolated from a human mixed lymphocyte reaction library based on its
CC   sequence similarity to genes encoding members of the interleukin-1
CC   (IL-1) receptor superfamily. The TANGO 191 and TANGO 195 nucleic acids
CC   and proteins can be used in screening assays, detection assays and
CC   pharmacogenomics. The nucleic acids can be used in gene therapy and for
CC   detecting mRNA or genetic lesions. In particular TANGO 191 modulators
CC   are useful in treatment of immune and inflammatory disorders, e.g.
CC   asthma, graft-versus-host disease, rheumatoid arthritis, cancer and
CC   autoimmune diseases, e.g. myasthenia gravis.
```



```
XX
SQ   Sequence 1118 BP; 227 A; 327 C; 301 G; 263 T; 0 other;


  Query Match             0.3%;  Score 20;  DB 21;  Length 1118;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db   1098 ttaatgaaaaaaaaaaaaaa 1117


RESULT  67
V75036
ID   V75036 standard; DNA; 1440 BP.
XX
AC   V75036;
XX
DT   16-MAR-1999  (first entry)
XX
DE   Staphylococcus aureus contig SEQ ID #725.
XX
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1021..1080
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
XX
PN   EP786519-A2.
XX
PD   30-JUL-1997.
XX
PF   07-JAN-1997;   97EP-0100117.
XX
PR   05-JAN-1996;   96US-0009861.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Fannon MR,  Kunsch CA;
PI   Rosen CA;
XX
```



```
DR   WPI; 1997-374922/35.
XX
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
XX
PS   Claim 1; Page 1628-1629; 3271pp; English.
XX
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
XX
SQ   Sequence 1440 BP; 519 A; 179 C; 241 G; 437 T; 64 other;


  Query Match             0.3%;  Score 20;  DB 18;  Length 1440;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    809 atttaatcaacaaagacaat 828
          ||||||||||||||||||||
Db   1405 atttaatcaacaaagacaat 1424


RESULT  68
T72662
ID   T72662 standard; cDNA; 1537 BP.
XX
AC   T72662;
XX
DT   15-SEP-1997  (first entry)
XX
DE   Human smooth muscle cell-derived migration factor cDNA.
XX
KW   Smooth muscle cell-derived migration factor; SDMF; atherosclerosis;
KW   restenosis; therapy; diagnosis; ss.
```

```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             94..1443
FT                   /*tag= a
FT                   /note= "Claim 1"
XX
PN   WO9719704-A1.
XX
PD   05-JUN-1997.
XX
PF   28-NOV-1995;    95WO-US15249.
XX
PR   28-NOV-1995;    95WO-US15249.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hurle MR,  McDonnell PC,  McNulty DE,  Rosen CA,  Siemens IR;
PI   Young PR,  Yue T;
XX
DR   WPI; 1997-310364/28.
DR   P-PSDB; W19720.
XX
PT   Human smooth muscle cell-derived migration factor polynucleotide -
PT   used to develop products for diagnosis and therapy, e.g. for
PT   atherosclerosis, restenosis and other blood vessel pathologies
XX
PS   Claim 1; Page 33-34; 57pp; English.
XX
CC   A cDNA clone (T72662), of which the coding sequence is claimed,
CC   codes for human smooth muscle cell-derived migration factor (SDMF)
CC   (W19720). It was isolated from a human foetal heart cDNA library
CC   in lambda ZAP using a murine p14 sequence (see also T72663) that had
CC   itself been identified using a rat SDMF tryptic peptide (W19721).
CC   Human SDMF nucleic acids can be used in methods of diagnosing
CC   conditions associated with SDMF protein, in the gene therapy of
CC   conditions related to insufficient SDMF protein function, in
CC   antisense form to modulate SDMF function for the treatment and
CC   prevention of atherosclerosis, restenosis or other blood vessel
CC   pathologies where smooth muscle cell migration is involved, to
CC   breed transgenic animals, and to produce SDMF recombinant
CC   polypeptides.
XX
SQ   Sequence 1537 BP; 312 A; 512 C; 423 G; 290 T; 0 other;

  Query Match             0.3%;  Score 20;  DB 18;  Length 1537;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db   1505 ttaatgaaaaaaaaaaaaa 1524


RESULT  69
X90357
ID   X90357 standard; cDNA; 1537 BP.
XX
AC   X90357;
XX
DT   27-SEP-1999  (first entry)
XX
DE   Human smooth muscle cell-derived migration factor encoding cDNA.
XX
KW   Smooth muscle cell-derived migration factor; SDMF; restenosis;
KW   atherosclerosis; blood vessel pathology; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             94..1443
FT                   /*tag= a
FT                   /product= "SDMF"
XX
PN   US5916758-A.
XX
PD   29-JUN-1999.
XX
PF   23-APR-1997;    97US-0839008.
XX
PR   08-SEP-1995;    95US-0003344.
PR   28-NOV-1995;    95US-0563697.
PR   23-APR-1997;    97US-0839008.
XX
PA   (HURL/) HURLE M R.
PA   (MCDO/) MCDONNELL P C.
PA   (MCNU/) MCNULTY D E.
PA   (ROSE/) ROSEN C A.
PA   (SIEM/) SIEMENS I R.
PA   (YOUN/) YOUNG P R.
PA   (YUET/) YUE T.
XX
PI   Hurle MR,  McDonnell PC,  McNulty DE,  Rosen CA,  Siemens IR;
PI   Young PR,  Yue T;
XX
DR   WPI; 1999-442764/37.
DR   P-PSDB; Y24473.
XX
PT   Screening for agents which modulate the activity of smooth muscle
```



```
PT   cell-derived migration factor
XX
PS   Example 1; Column 17-20; 21pp; English.
XX
CC   The present invention describes a method of screening for agents which
CC   modulate the activity of human smooth muscle cell-derived migration
CC   factor (SDMF). SDMF causes smooth muscle cells (SMCs) to migrate into
CC   the intimal layer of blood vessel walls. The method may be used to
CC   screen for agents which modulate (especially inhibit) the activity of
CC   human SDMF by interfering with the binding of the SDMF to binding
CC   partners. Modulators which may modulate SDMF activity include cell
CC   surface receptors, antibodies, binding proteins, peptides,
CC   oligonucleotides and small organic molecules such as peptidomimetics.
CC   The modulators may be used in the prevention and treatment of various
CC   pathologies arising from over-/inappropriate-production of SDMF. Such
CC   pathologies include restenosis and atherosclerosis or other blood
CC   vessel pathologies where SMC migration is involved. The present
CC   sequence encodes human SDMF.
XX
SQ   Sequence 1537 BP; 312 A; 512 C; 423 G; 290 T; 0 other;

  Query Match             0.3%;  Score 20;  DB 20;  Length 1537;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db   1505 ttaatgaaaaaaaaaaaaa 1524


RESULT  70
V34250
ID   V34250 standard; DNA; 1969 BP.
XX
AC   V34250;
XX
DT   28-JAN-1999  (first entry)
XX
DE   Human secreted protein gene 35 clone HTXAK60.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
```



```
XX
PN   WO9839446-A2.
XX
PD   11-SEP-1998.
XX
PF   06-MAR-1998;    98WO-US04492.
XX
PR   07-MAR-1997;    97US-0038621.
PR   07-MAR-1997;    97US-0040161.
PR   07-MAR-1997;    97US-0040162.
PR   07-MAR-1997;    97US-0040163.
PR   07-MAR-1997;    97US-0040333.
PR   07-MAR-1997;    97US-0040334.
PR   07-MAR-1997;    97US-0040336.
PR   07-MAR-1997;    97US-0040626.
PR   11-APR-1997;    97US-0043311.
PR   11-APR-1997;    97US-0043312.
PR   11-APR-1997;    97US-0043313.
PR   11-APR-1997;    97US-0043314.
PR   11-APR-1997;    97US-0043315.
PR   11-APR-1997;    97US-0043568.
PR   11-APR-1997;    97US-0043569.
PR   11-APR-1997;    97US-0043576.
PR   11-APR-1997;    97US-0043578.
PR   11-APR-1997;    97US-0043580.
PR   11-APR-1997;    97US-0043669.
PR   11-APR-1997;    97US-0043670.
PR   11-APR-1997;    97US-0043671.
PR   11-APR-1997;    97US-0043672.
PR   11-APR-1997;    97US-0043674.
PR   23-MAY-1997;    97US-0047492.
PR   23-MAY-1997;    97US-0047500.
PR   23-MAY-1997;    97US-0047501.
PR   23-MAY-1997;    97US-0047502.
PR   23-MAY-1997;    97US-0047503.
PR   23-MAY-1997;    97US-0047581.
PR   23-MAY-1997;    97US-0047582.
PR   23-MAY-1997;    97US-0047583.
PR   23-MAY-1997;    97US-0047584.
PR   23-MAY-1997;    97US-0047585.
PR   23-MAY-1997;    97US-0047586.
PR   23-MAY-1997;    97US-0047587.
PR   23-MAY-1997;    97US-0047588.
PR   23-MAY-1997;    97US-0047589.
PR   23-MAY-1997;    97US-0047590.
PR   23-MAY-1997;    97US-0047592.
PR   23-MAY-1997;    97US-0047593.
PR   23-MAY-1997;    97US-0047594.
PR   23-MAY-1997;    97US-0047595.
PR   23-MAY-1997;    97US-0047596.
PR   23-MAY-1997;    97US-0047597.
```

```
PR   23-MAY-1997;   97US-0047598.
PR   23-MAY-1997;   97US-0047599.
PR   23-MAY-1997;   97US-0047600.
PR   23-MAY-1997;   97US-0047601.
PR   23-MAY-1997;   97US-0047612.
PR   23-MAY-1997;   97US-0047613.
PR   23-MAY-1997;   97US-0047614.
PR   23-MAY-1997;   97US-0047615.
PR   23-MAY-1997;   97US-0047617.
PR   23-MAY-1997;   97US-0047618.
PR   23-MAY-1997;   97US-0047632.
PR   23-MAY-1997;   97US-0047633.
PR   06-JUN-1997;   97US-0048964.
PR   06-JUN-1997;   97US-0048974.
PR   22-AUG-1997;   97US-0056630.
PR   22-AUG-1997;   97US-0056631.
PR   22-AUG-1997;   97US-0056632.
PR   22-AUG-1997;   97US-0056636.
PR   22-AUG-1997;   97US-0056637.
PR   22-AUG-1997;   97US-0056662.
PR   22-AUG-1997;   97US-0056664.
PR   22-AUG-1997;   97US-0056845.
PR   22-AUG-1997;   97US-0056862.
PR   22-AUG-1997;   97US-0056864.
PR   22-AUG-1997;   97US-0056872.
PR   22-AUG-1997;   97US-0056874.
PR   22-AUG-1997;   97US-0056875.
PR   22-AUG-1997;   97US-0056876.
PR   22-AUG-1997;   97US-0056877.
PR   22-AUG-1997;   97US-0056878.
PR   22-AUG-1997;   97US-0056879.
PR   22-AUG-1997;   97US-0056880.
PR   22-AUG-1997;   97US-0056881.
PR   22-AUG-1997;   97US-0056882.
PR   22-AUG-1997;   97US-0056884.
PR   22-AUG-1997;   97US-0056886.
PR   22-AUG-1997;   97US-0056887.
PR   22-AUG-1997;   97US-0056888.
PR   22-AUG-1997;   97US-0056889.
PR   22-AUG-1997;   97US-0056892.
PR   22-AUG-1997;   97US-0056893.
PR   22-AUG-1997;   97US-0056894.
PR   22-AUG-1997;   97US-0056903.
PR   22-AUG-1997;   97US-0056908.
PR   22-AUG-1997;   97US-0056909.
PR   22-AUG-1997;   97US-0056910.
PR   22-AUG-1997;   97US-0056911.
PR   05-SEP-1997;   97US-0057650.
PR   05-SEP-1997;   97US-0057761.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
```



```
XX
PI   Bednarik DP,  Brewer LA,  Carter KC,  Duan R,  Ebner R,  Endress GA;
PI   Feng P,  Ferrie AM,  Fischer CL,  Graves KA,  Greene JM,  Hu JS;
PI   Kyaw H,  Lafleur DW,  Li Y,  Moore PA,  Ni J,  Olsen HS,  Rosen CA;
PI   Ruben SM,  Shi Y,  Soppet DR,  Young PE,  Yu GL,  Zeng Z;
XX
DR   WPI; 1998-609887/51.
DR   P-PSDB; W75153.
XX
PT   New isolated human genes and the secreted polypeptides they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 250-251; 447pp; English.
XX
CC   This sequence represents a nucleic acid molecule which encodes a secreted
CC   human protein.  The gene number, and the clone it is derived from, are
CC   detailed in the descriptor line.. The gene can be used to generate fusion
CC   proteins by linking to the gene to a human immunoglobulin Fc portion
CC   (e.g. V34145) for increasing the stability of the fused protein as
CC   compared to the human protein only.
CC   The invention relates to 70 novel genes and their fragments (nucleic acid
CC   sequences: V34154-V34276; amino acid sequences W75057-W75179) which
CC   are useful for preventing, treating or ameliorating  medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 70 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V34154 for described
CC   uses).
XX
SQ   Sequence 1969 BP; 448 A; 532 C; 540 G; 442 T; 7 other;


  Query Match             0.3%;  Score 20;  DB 19;  Length 1969;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    1909 ttaatgaaaaaaaaaaaaaa 1928


RESULT  71
V38661
ID   V38661 standard; DNA; 2187 BP.
XX
AC   V38661;
XX
DT   27-OCT-1998  (first entry)
XX
```



```
DE   Mus musculus SOCS3 gene.
XX
KW   SOCS; suppressor of cytokine signalling; PCR primer;
KW   autoimmune disease; diagnosis; cancer; treatment;
KW   cytokine mediated cellular responsiveness; hyperimmunity;
KW   immunosuppression; allergies; hypertension; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             18..695
FT                   /*tag= a
FT                   /product= SOCS3 protein
XX
PN   WO9820023-A1.
XX
PD   14-MAY-1998.
XX
PF   31-OCT-1997;   97WO-AU00729.
XX
PR   14-FEB-1997;   97AU-0005117.
PR   01-NOV-1996;   96AU-0003384.
XX
PA   (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI   Alexander WS,  Hilton DJ,  Metcalf D,  Nicholson SE;
PI   Nicola NA,  Richardson RT,  Starr R,  Viney EM,  Willson TA;
XX
DR   WPI; 1998-286854/25.
DR   P-PSDB; W62615.
XX
PT   Suppressor of cytokine signalling proteins - useful to treat
PT   disease, injury or abnormality involving cytokine mediated cellular
PT   responsiveness e.g. hyperimmunity, immunosuppression, allergies and
PT   hypertension
XX
PS   Claim 14; Page 113-114; 325pp; English.
XX
CC   The sequence is that of a gene encoding a suppressor of cytokine
CC   signalling protein (SOCS). SOCS can be used to screen for naturally
CC   occurring antibodies to SOCS, which may occur, e.g. in some autoimmune
CC   diseases. Alternatively, specific antibodies can be used to
CC   screen for SOCS, which is useful as a knowledge of SOCS levels
CC   may be important for the diagnosis of certain cancers. Soluble
CC   SOCS polypeptides can be used to treat disease, injury or
CC   abnormality involving cytokine mediated cellular responsiveness,
CC   e.g. hyperimmunity, immunosuppression, allergies and hypertension.
XX
SQ   Sequence 2187 BP; 501 A; 632 C; 600 G; 454 T; 0 other;
```



```
  Query Match             0.3%;  Score 20;  DB 19;  Length 2187;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2156 ttaatgaaaaaaaaaaaaaa 2175


RESULT  72
X90601
ID   X90601 standard; DNA; 2201 BP.
XX
AC   X90601;
XX
DT   01-OCT-1999  (first entry)
XX
DE   Histidine kinase encoding polynucleotide sequence.
XX
KW   Histidine kinase; Two component signal transduction system; TCSTS;
KW   virulence; bacterial pathogenesis; autophosphorylate; immunisation;
KW   antibacterial agent; cognate response regulator; screening;
KW   Helicobacter pylori; gastric ulcer; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   CDS             744..2099
FT                   /*tag= a
FT                   /product= "Histidine kinase protein"
FT                   /note= "Has sequence homology to KinA from Lactococcus
FT                   lactis cremoris"
XX
PN   WO9936508-A1.
XX
PD   22-JUL-1999.
XX
PF   12-JAN-1999;   99WO-US00610.
XX
PR   13-JAN-1998;   98US-0006627.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Debouck C,  Jaworski DD,  Mooney JL,  Shilling LK;
PI   Throup JP,  Wallis NG,  Wang M,  Zhong Y;
XX
DR   WPI; 1999-444390/37.
DR   P-PSDB; Y28601.
XX
PT   Novel histidine kinase polynucleotides and polypeptides used to
PT   screen for antibacterial compounds
```

```
XX
PS   Claim 6; Page 5-6; 43pp; English.
XX
CC   The present sequence is a histidine kinase encoding novel polynucleotide
CC   isolated from Staphylococcus aureus. The sequence was obtained from a
CC   library of chromosomal DNA clones of Staphylococcus aureus in E.coli.
CC   Histidine kinase is a component of the two component signal transduction
CC   system(TCSTS) which is involved in bacterial pathogenesis and virulence.
CC   It undergoes autophosphorylation at a histidine residue and phosphate
CC   group is transferred to the cognate response regulator. The nucleotide
CC   sequence can be used for diagnosis and staging of diseases, and as
CC   reagents for screening genetic mutations. The polynucleotide can also
CC   be used for therapeutic or prophylactic purposes and in particular for
CC   genetic immunisation. Polypeptides derived from this sequence can be
CC   used to produce antibodies and to identify agonists and antagonists
CC   which are used to prevent, inhibit or treat diseases, particularly
CC   Helicobacter pylori infections, such as gastric ulcers, gastrointestinal
CC   carcinoma, and gastritis. The histidine kinase products can be used to
CC   screen new antibacterial drugs effective against resistant S.aureus
CC   strains.
XX
SQ   Sequence 2201 BP; 863 A; 290 C; 372 G; 676 T; 0 other;


  Query Match             0.3%;  Score 20;  DB 20;  Length 2201;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     809 atttaatcaacaaagacaat 828
           ||||||||||||||||||||
Db    1430 atttaatcaacaaagacaat 1449


RESULT  73
Q73786
ID   Q73786 standard; cDNA; 2309 BP.
XX
AC   Q73786;
XX
DT   26-JUN-1995  (first entry)
XX
DE   Partial PCR fragment of PTP-S31D.
XX
KW   Protein tyrosine phosphatase; cancer; diabetes; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             65..1035
FT                   /*tag= a
FT   misc_feature    1766
```



```
FT                   /*tag= b
FT                   /note= "unknown base"
XX
PN   WO9421800-A.
XX
PD   29-SEP-1994.
XX
PF   23-MAR-1994;    94WO-EP00909.
XX
PR   23-MAR-1993;    93US-0036210.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Moller KB,  Moller NPH,  Ullrich A;
XX
DR   WPI; 1994-317020/39.
DR   P-PSDB; R60877.
XX
PT   New protein tyrosine phosphatase PTP-S31 - is used to develop
PT   prods. for treating or preventing disease associated with
PT   abnormal PTP-S31, e.g. cancer or diabetes
XX
PS   Claim 6; Fig 6; 116pp; English.
XX
CC   The sequence is that of a partial fragment of a new protein tyrosine
CC   phospatase PTP-S31D that was isolated from RNA from human skeletal
CC   muscle.  Such DNA can be used to treat cancer and diabetes.
CC   See also Q73782-8.
XX
SQ   Sequence 2309 BP; 770 A; 384 C; 435 G; 719 T; 1 other;


  Query Match             0.3%;  Score 20;  DB 15;  Length 2309;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2274 ttaatgaaaaaaaaaaaaa 2293


RESULT  74
V34188
ID   V34188 standard; DNA; 2378 BP.
XX
AC   V34188;
XX
DT   28-JAN-1999  (first entry)
XX
DE   Human secreted protein gene 35 clone HTXAK60.
XX
```



```
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9839446-A2.
XX
PD   11-SEP-1998.
XX
PF   06-MAR-1998;    98WO-US04492.
XX
PR   07-MAR-1997;    97US-0038621.
PR   07-MAR-1997;    97US-0040161.
PR   07-MAR-1997;    97US-0040162.
PR   07-MAR-1997;    97US-0040163.
PR   07-MAR-1997;    97US-0040333.
PR   07-MAR-1997;    97US-0040334.
PR   07-MAR-1997;    97US-0040336.
PR   07-MAR-1997;    97US-0040626.
PR   11-APR-1997;    97US-0043311.
PR   11-APR-1997;    97US-0043312.
PR   11-APR-1997;    97US-0043313.
PR   11-APR-1997;    97US-0043314.
PR   11-APR-1997;    97US-0043315.
PR   11-APR-1997;    97US-0043568.
PR   11-APR-1997;    97US-0043569.
PR   11-APR-1997;    97US-0043576.
PR   11-APR-1997;    97US-0043578.
PR   11-APR-1997;    97US-0043580.
PR   11-APR-1997;    97US-0043669.
PR   11-APR-1997;    97US-0043670.
PR   11-APR-1997;    97US-0043671.
PR   11-APR-1997;    97US-0043672.
PR   11-APR-1997;    97US-0043674.
PR   23-MAY-1997;    97US-0047492.
PR   23-MAY-1997;    97US-0047500.
PR   23-MAY-1997;    97US-0047501.
PR   23-MAY-1997;    97US-0047502.
PR   23-MAY-1997;    97US-0047503.
PR   23-MAY-1997;    97US-0047581.
PR   23-MAY-1997;    97US-0047582.
PR   23-MAY-1997;    97US-0047583.
PR   23-MAY-1997;    97US-0047584.
PR   23-MAY-1997;    97US-0047585.
PR   23-MAY-1997;    97US-0047586.
```



```
PR   23-MAY-1997;    97US-0047587.
PR   23-MAY-1997;    97US-0047588.
PR   23-MAY-1997;    97US-0047589.
PR   23-MAY-1997;    97US-0047590.
PR   23-MAY-1997;    97US-0047592.
PR   23-MAY-1997;    97US-0047593.
PR   23-MAY-1997;    97US-0047594.
PR   23-MAY-1997;    97US-0047595.
PR   23-MAY-1997;    97US-0047596.
PR   23-MAY-1997;    97US-0047597.
PR   23-MAY-1997;    97US-0047598.
PR   23-MAY-1997;    97US-0047599.
PR   23-MAY-1997;    97US-0047600.
PR   23-MAY-1997;    97US-0047601.
PR   23-MAY-1997;    97US-0047612.
PR   23-MAY-1997;    97US-0047613.
PR   23-MAY-1997;    97US-0047614.
PR   23-MAY-1997;    97US-0047615.
PR   23-MAY-1997;    97US-0047617.
PR   23-MAY-1997;    97US-0047618.
PR   23-MAY-1997;    97US-0047632.
PR   23-MAY-1997;    97US-0047633.
PR   06-JUN-1997;    97US-0048964.
PR   06-JUN-1997;    97US-0048974.
PR   22-AUG-1997;    97US-0056630.
PR   22-AUG-1997;    97US-0056631.
PR   22-AUG-1997;    97US-0056632.
PR   22-AUG-1997;    97US-0056636.
PR   22-AUG-1997;    97US-0056637.
PR   22-AUG-1997;    97US-0056662.
PR   22-AUG-1997;    97US-0056664.
PR   22-AUG-1997;    97US-0056845.
PR   22-AUG-1997;    97US-0056862.
PR   22-AUG-1997;    97US-0056864.
PR   22-AUG-1997;    97US-0056872.
PR   22-AUG-1997;    97US-0056874.
PR   22-AUG-1997;    97US-0056875.
PR   22-AUG-1997;    97US-0056876.
PR   22-AUG-1997;    97US-0056877.
PR   22-AUG-1997;    97US-0056878.
PR   22-AUG-1997;    97US-0056879.
PR   22-AUG-1997;    97US-0056880.
PR   22-AUG-1997;    97US-0056881.
PR   22-AUG-1997;    97US-0056882.
PR   22-AUG-1997;    97US-0056884.
PR   22-AUG-1997;    97US-0056886.
PR   22-AUG-1997;    97US-0056887.
PR   22-AUG-1997;    97US-0056888.
PR   22-AUG-1997;    97US-0056889.
PR   22-AUG-1997;    97US-0056892.
PR   22-AUG-1997;    97US-0056893.
```

```
PR   22-AUG-1997;   97US-0056894.
PR   22-AUG-1997;   97US-0056903.
PR   22-AUG-1997;   97US-0056908.
PR   22-AUG-1997;   97US-0056909.
PR   22-AUG-1997;   97US-0056910.
PR   22-AUG-1997;   97US-0056911.
PR   05-SEP-1997;   97US-0057650.
PR   05-SEP-1997;   97US-0057761.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bednarik DP,  Brewer LA,  Carter KC,  Duan R,  Ebner R,  Endress GA;
PI   Feng P,  Ferrie AM,  Fischer CL,  Graves KA,  Greene JM,  Hu JS;
PI   Kyaw H,  Lafleur DW,  Li Y,  Moore PA,  Ni J,  Olsen HS,  Rosen CA;
PI   Ruben SM,  Shi Y,  Soppet DR,  Young PE,  Yu GL,  Zeng Z;
XX
DR   WPI; 1998-609887/51.
DR   P-PSDB; W75091.
XX
PT   New isolated human genes and the secreted polypeptides they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 190-191; 447pp; English.
XX
CC   This sequence representa a nucleic acid molecule which encodes a secreted
CC   human protein.  The gene number, and the clone it is derived from, are
CC   detailed in the descriptor line.  The gene can be used to generate fusion
CC   proteins by linking to the gene to a human immunoglobulin Fc portion
CC   (e.g. V34145) for increasing the stability of the fused protein as
CC   compared to the human protein only.
CC   The invention relates to 70 novel genes and their fragments (nucleic acid
CC   sequences: V34154-V34276; amino acid sequences W75057-W75179) which
CC   are useful for preventing, treating or ameliorating  medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 70 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V34154 for described
CC   uses).
XX
SQ   Sequence 2378 BP; 518 A; 690 C; 662 G; 506 T; 2 other;


  Query Match           0.3%;  Score 20;  DB 19;  Length 2378;
  Best Local Similarity  100.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2355 ttaatgaaaaaaaaaaaaa 2374
```



```
RESULT  75
Q20576/c
ID   Q20576 standard; cDNA; 2529 BP.
XX
AC   Q20576;
XX
DT   05-MAY-1992  (first entry)
XX
DE   Transforming Growth Factor beta 3 coding sequence.
XX
KW   TGF-beta 3; homodimer; ss.
XX
OS   Synthetic.
XX
PN   WO9200318-A.
XX
PD   09-JAN-1992.
XX
PF   25-JUN-1991;   91WO-US04541.
XX
PR   25-JUN-1990;   90US-0543348.
XX
PA   (ONCO-) ONCOGENE SCI INC.
XX
PI   Iwata KK,  Foulkes JG,  Tendijke P,  Haley JD;
XX
DR   WPI; 1992-041510/05.
DR   P-PSDB; R20621.
XX
PT   Transforming growth factor beta 3 proteins, precursors and
PT   mutants - obtd. from polypeptide and antibodies, with optimal
PT   therapeutic use due to genetic manipulation of coding sequence
XX
PS   Example 2; Fig 1; 107pp; English.
XX
CC   This 2529bp full-length TGF-beta3 gene sequence was obtained from
CC   three shorter overlapping clones derived from human placental, human
CC   umbilical cord and A673 cells cDNA libraries, respectively. The
CC   predicted amino acid sequence of the gene encoding TGF-beta3 shows
CC   extensive homology to  TGF-beta 1 and beta 2.
CC   See also Q22229 and R20622.
XX
SQ   Sequence 2529 BP; 617 A; 670 C; 661 G; 581 T; 0 other;


  Query Match           0.3%;  Score 20;  DB 13;  Length 2529;
  Best Local Similarity  100.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy     389 aaaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db    1981 AAAAAAAAAAAAAATGCTTG 1962


RESULT  76
Q06845/c
ID   Q06845 standard; cDNA; 2530 BP.
XX
AC   Q06845;
XX
DT   05-MAR-1991  (first entry)
XX
DE   Sequence encoding tumour growth inhibitor.
XX
KW   TGI; carcinoma; melanoma; leukaemia; arteriosclerosis; inflammation;
KW   psoriasis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             263..1501
FT                   /*tag=  a
FT   polyA_signal    2506..2511
FT                   /*tag=  b
FT   polyA_site      2529
FT                   /*tag=  c
XX
PN   WO9014360-A.
XX
PD   29-NOV-1990.
XX
PF   17-MAY-1990;   90WO-US02753.
XX
PR   17-MAY-1989;   89US-0353410.
PR   19-APR-1985;   85US-0725003.
PR   07-APR-1986;   86US-0847931.
PR   20-OCT-1986;   86US-0992121.
PR   20-OCT-1987;   87US-0111022.
PR   20-APR-1988;   88US-0183224.
XX
PA   (ONCO-) ONCOGENE SCI INC.
XX
PI   Iwata KK,  Stephenson JR,  Tendijke P,  Franco R,  Gold LI,  Foulkes JG;
XX
DR   WPI; 1990-375949/50.
DR   P-PSDB; R08264.
XX
PT   Tissue-derived tumour growth inhibitors - comprise specified
PT   protein sequences used to detect, and treat tumours, burns and
PT   wounds.
```



```
XX
PS   Claim 10; Fig 41; 190pp; English.
XX
CC   The plasmid was isolated from a human cDNA library prepd. from the
CC   DNA of a chronic myelocytic leukaemia cell line (K562).  The
CC   sequence can be used to produce the tumour growth inhibitor (TGI)
CC   by recombinant techniques.  The protein may also be isolated from
CC   human umbilical cord and placental tissues.  It can be used to
CC   inhibit tumour cell growth, to treat burns, to facilitate the
CC   healing of wounds or to treat proliferative disorders.  The
CC   protein and Abs raised to it can be used for detection and typing
CC   of tumours.  The Abs can also be used to inhibit the activity of
CC   the TGI.
XX
SQ   Sequence 2530 BP; 619 A; 671 C; 659 G; 581 T; 0 other;


  Query Match           0.3%;  Score 20;  DB 11;  Length 2530;
  Best Local Similarity  100.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db    1981 AAAAAAAAAAAAAATGCTTG 1962


RESULT  77
Q39221
ID   Q39221 standard; DNA; 2734 BP.
XX
AC   Q39221;
XX
DT   28-JUL-1993  (first entry)
XX
DE   Encodes human epidermal transglutaminase.
XX
KW   skin; epidermis; aminotransferase enzymes.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             81..2534
FT                   /*tag=  a
XX
PN   JP05056785-A.
XX
PD   09-MAR-1993.
XX
PF   28-MAR-1991;   91JP-0065034.
XX
PR   28-MAR-1991;   91JP-0065034.
```

```
XX
PA   (ADOB-) ADOBANSUTOSUKIN RES KENKYUSHO KK.
XX
DR   WPI; 1993-120381/15.
DR   P-PSDB; R34942.
XX
PT   New DNA sequence coding human epidermis trans:glutaminase - is
PT   used to prepare the trans:glutaminase by culturing host cells
PT   contg. the recombinant vector of the sequence
XX
PS   Claim 1; Page 5; 10pp; Japanese.
XX
CC   This sequence encodes human epidermal transglutaminase. Primers
CC   Q39222,3 were used in the isolation of its cDNA.
XX
SQ   Sequence 2734 BP; 610 A; 781 C; 806 G; 537 T; 0 other;
```

```
  Query Match             0.3%;  Score 20;  DB 14;  Length 2734;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2704 ttaatgaaaaaaaaaaaaa 2723
```

```
RESULT  78
Q47376/c
ID   Q47376 standard; mRNA; 2734 BP.
XX
AC   Q47376;
XX
DT   14-FEB-1994  (first entry)
XX
DE   hETG mRNA.
XX
KW   Human; epidermal transglutaminase; hETG; probe; detection; tumours;
KW   keratinisation; ss.
XX
OS   Homo sapiens.
XX
PN   JP05184396-A.
XX
PD   27-JUL-1993.
XX
PF   17-JAN-1992;   92JP-0006358.
XX
PR   17-JAN-1992;   92JP-0006358.
XX
PA   (ADSK-) ADVANCED SKIN RES KENKYUSHO KK.
```



```
XX
DR   WPI; 1993-269049/34.
XX
PT   Diagnosis method for the detection of keratinisation of a cell -
PT   to detect the progess of tumours
XX
PS   Disclosure; Page 5-6; 7pp; Japanese.
XX
CC   This sequence represents the human epidermal transglutaminase (hETG)
CC   mRNA.  A fragment of this sequence was used as a probe to detect
CC   keratinisation of cells.  The progression of tumours could, therefore
CC   be monitored.
XX
SQ   Sequence 2734 BP; 537 A; 806 C; 781 G; 610 U; 0 other;
```

```
  Query Match             0.3%;  Score 20;  DB 14;  Length 2734;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      31 TTAATGAAAAAAAAAAAAA 12
```

```
RESULT  79
A09042
ID   A09042 standard; cDNA; 2894 BP.
XX
AC   A09042;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Full-length human TANGO 195 cDNA (clone T195AthLa170f10).
XX
KW   TANGO 195; transmembrane protein; modulator; proliferation;
KW   differentiation; cell survival; CD2; Ig superfamily;
KW   gene therapy; anti-inflammatory; anti-allergic; cytostatic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             25..882
FT                   /*tag= a
FT   sig_peptide     25..87
FT                   /*tag= b
FT   mat_peptide     88..879
FT                   /*tag= c
XX
PN   WO200018800-A1.
XX
```



```
PD   06-APR-2000.
XX
PF   30-SEP-1999;   99WO-US22818.
XX
PR   30-SEP-1998;   98US-0163523.
XX
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI   Busfield SJ;
XX
DR   WPI; 2000-293110/25.
DR   P-PSDB; Y92183.
XX
PT   New cDNA molecules encoding human TANGO 191 and TANGO 195 proteins,
PT   useful for modulating cellular proliferation, differentiation or cell
PT   survival
XX
PS   Disclosure; Fig 9; 122pp; English.
XX
CC   Human TANGO 191 and 195 are transmembrane proteins, useful as modulating
CC   agents in a regulating a variety of cellular processes, e.g. cellular
CC   proliferation, differentiation and survival.  Human TANGO 195 is a type I
CC   transmembrane protein of the CD2 subgroup of the Ig superfamily.  It may
CC   play a role in B cell leukaemia, immune response and autoimmune
CC   disorders, e.g. arthritis.  The DNA sequence maps to human chromosome
CC   1q21.  The TANGO 191 and TANGO 195 nucleic acids
CC   and proteins can be used in screening assays, detection assays and
CC   pharmacogenomics.  The nucleic acids can be used in gene therapy and for
CC   detecting mRNA or genetic lesions.  In particular TANGO 191 modulators
CC   are useful in treatment of immune and inflammatory disorders, e.g.
CC   asthma, graft-versus-host disease, rheumatoid arthritis, cancer and
CC   autoimmune diseases, e.g. myasthenia gravis.
XX
SQ   Sequence 2894 BP; 733 A; 818 C; 682 G; 661 T; 0 other;
```

```
  Query Match             0.3%;  Score 20;  DB 21;  Length 2894;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2865 ttaatgaaaaaaaaaaaaa 2884
```

```
RESULT  80
V27941/c
ID   V27941 standard; cDNA; 14796 BP.
XX
AC   V27941;
XX
```



```
DT   25-SEP-1998  (first entry)
XX
DE   Survivin gene.
XX
KW   survivin; apoptosis; cellular apoptosis; transplantation; ss;
KW   motor neuron degenerative disease; HIV infection; immunosuppression;
KW   gastrointestinal perturbations; cardiovascular disorder.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2811..12044
FT                   /*tag= a
FT                   /product= Survivin
FT   exon            2811..2921
FT                   /*tag= b
FT                   /number= 1
FT   intron          2922..3173
FT                   /*tag= c
FT                   /number= 1
FT   exon            3174..3283
FT                   /*tag= d
FT                   /number= 2
FT   intron          3284..5157
FT                   /*tag= e
FT                   /number= 2
FT   exon            5158..5275
FT                   /*tag= f
FT                   /number= 3
FT   intron          5276..11954
FT                   /*tag= g
FT                   /number= 3
FT   exon            11955..12044
FT                   /*tag= h
FT                   /number= 4
XX
PN   WO9822589-A2.
XX
PD   28-MAY-1998.
XX
PF   20-NOV-1997;   97WO-US21880.
XX
PR   20-NOV-1997;   97US-0975080.
PR   20-NOV-1996;   96US-0031435.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Altieri DC;
XX
DR   WPI; 1998-312475/27.
DR   P-PSDB; W6143S9.
```

XX
PT Modulating apoptosis by controlling the Survivin gene - useful for
PT treating transplant rejection, degenerative disorders and tumours
XX
PS Disclosure; Fig 10; 108pp; English.
XX
CC The survivin gene can be used to control apoptosis through modification
CC of the gene. Survivin peptides can be used to inhibit cellular
CC apoptosis, e.g. for enhancing the viability of organs and tissues prior
CC to their transplantation, for preserving the growth of cells in culture
CC or for treating conditions involving abnormal apoptosis,
CC e.g. degenerative diseases such as motor neuron degenerative diseases,
CC HIV infection, dermatological effects of ageing, disorders and diseases
CC such as immunosuppression, gastrointestinal perturbations, cardiovascular
CC disorders, apoptosis related to reperfusion damage, rejection of tissue
CC transplantation and Alzheimer's disease. Agents which block Survivin
CC activity can be used to treat e.g. tumours.
XX
SQ Sequence 14796 BP; 3482 A; 3488 C; 3888 G; 3938 T; 0 other;

Query Match 0.3%; Score 20; DB 19; Length 14796;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 388 gaaaaaaaaaaaaaatgctt 407
||||||||||||||||||||
Db 5389 GAAAAAAAAAAAAATGCTT 5370

RESULT 81
A09888/c
ID A09888 standard; DNA; 25871 BP.
XX
AC A09888;
XX
DT 05-JUL-2000 (first entry)
XX
DE Human genomic OCTN2 sequence.
XX
KW Organic cation transportation; human; carnitine transporter protein;
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
KW juvenile visceral steatosis; ds.
XX
OS Homo sapiens.
XX
PN WO200014210-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-JP04853.

XX
PR 07-SEP-1998; 98JP-0252683.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Nezu J, Oku A;
XX
DR WPI; 2000-256966/22.
XX
PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
PT cation transporter, useful as diagnostic tool
XX
PS Claim 1; Page 60-93; 106pp; Japanese.
XX
CC This sequence represents the genomic DNA encoding the human carnitine
CC transporter protein OCTN2 (Y83929). The complete genomic sequence
CC contains 9 introns dispersed throughout the gene. The gene can be used
CC as a target for diagnosis of systemic carnitine deficiency by detecting
CC the presence of mutations in the sequence, especially seen in the disease
CC juvenile visceral steatosis (jvs) . The wild type OCTN2 gene can be used
CC in the gene therapy of the disease state.
XX
SQ Sequence 25871 BP; 6356 A; 6094 C; 6408 G; 7013 T; 0 other;

Query Match 0.3%; Score 20; DB 21; Length 25871;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1876 tctaaaaaagacacaaaaca 1895
||||||||||||||||||||
Db 1805 TCTAAAAAAGACACAAAACA 1786

RESULT 82
V73526
ID V73526 standard; DNA; 19 BP.
XX
AC V73526;
XX
DT 22-MAR-2000 (first entry)
XX
DE H. pylori cag A primer cagSF.
XX
KW PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS Helicobacter pylori.

XX
PN WO9816658-A2.
XX
PD 23-APR-1998.
XX
PF 10-OCT-1997; 97WO-EP05614.
XX
PR 16-OCT-1996; 96EP-0870131.
PR 09-SEP-1997; 97EP-0870133.
XX
PA (INNO-) INNOGENETICS NV.
PA (DDLD-) DDL BV.
XX
PI Quint W, Van Doorn L;
XX
DR WPI; 1998-251300/22.
XX
PT Method for detecting and/or typing Helicobacter pylori strains -
PT comprises use of primers and probes based on vacA and cagA gene
XX
PS Claim 2: Page 44; 122pp; English.
XX
CC This invention describes a novel method for the detection and/or typing
CC of Helicobacter pylori strains present in a sample using PCR primers and
CC probes to detect regions of the vacuolating toxin (vacA) gene and other
CC virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC gene. The method allows the typing and allele-specific detection of a
CC strain according to the VDG alleles present in that particular H. pylori
CC strain. The virulence determinant genes are the genetic elements involved
CC in enabling, determining, and marking the infectivity and/or
CC pathogenicity of the H. pylori strain. The method provides a way of
CC detecting H. pylori strains in a sample with respect to the development
CC of chronic active gastritis, gastric and duodenal ulcers, gastric
CC adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC determining eradication therapy. V73508-V73546 represent PCR primers
CC and probes used in the detection of the H. pylori vacA and cagA genes.
CC The primers and probes are used especially to detect the vacA S regions
CC S1a/b/c and S2 and the M regions M1 and M2 which are represented in
CC V73547-V73785.
XX
SQ Sequence 19 BP; 10 A; 7 C; 2 G; 0 U; 0 other;

Query Match 0.3%; Score 19; DB 19; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 556 caacaaccacaaaccgaag 574
|||||||||||||||||||
Db 1 caacaaccacaaaccgaag 19

RESULT 83
V25348
ID V25348 standard; DNA; 28 BP.
XX
AC V25348;
XX
DT 11-JUN-1998 (first entry)
XX
DE Primer for H.pylori murI gene.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW PCR primer; ss.
XX
OS Synthetic.
OS Helicobacter pylori.
XX
PN WO9737044-A1.
XX
PD 09-OCT-1997.
XX
PF 27-MAR-1997; 97WO-US05223.
XX
PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
PA (ASTR) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR WPI; 1997-503122/46.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Example; Page 121; 1145pp; English.
XX
CC This sequence represents a primer for the H.pylori murI gene.
CC The amplified sequence was used to compare homology of the coding
CC sequences of the invention with other known proteins. The protein encoded
CC by the DNA of the invention may be used in a vaccine to prevent or treat
CC H.pylori infection or to identify H.pylori polypeptide binding compounds,
CC useful as potential H.pylori life cycle activators or inhibitors. The DNA
CC and probes derived from it may be used for the identification of H.pylori
CC in a sample and the diagnosis of H.pylori infection. Nucleic acid

```
CC   sequences complementary to the DNA act as antisense sequences and can be
CC   used to prevent the translation of H.pylori mRNA. Antibodies against the
CC   protein can be used in immunoassays to evaluate the abundance and
CC   distribution of H.pylori-specific antigens. The genomic sequence of
CC   H.pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H.pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H.pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
XX
SQ   Sequence 28 BP; 9 A; 5 C; 3 G; 11 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 18;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5050 ttacaatttgagccattct 5068
           |||||||||||||||||||
Db      10 ttacaatttgagccattct 28


RESULT  84
V25347/c
ID   V25347 standard; DNA; 29 BP.
XX
AC   V25347;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Primer for H.pylori murI gene.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   PCR primer; ss.
XX
OS   Synthetic.
OS   Helicobacter pylori.
XX
PN   WO9737044-A1.
XX
PD   09-OCT-1997.
XX
PF   27-MAR-1997;   97WO-US05223.
XX
```



```
PR   06-DEC-1996;   96US-0761318.
PR   29-MAR-1996;   96US-0625811.
PR   02-APR-1996;   96US-0758731.
PR   25-OCT-1996;   96US-0736905.
PR   28-OCT-1996;   96US-0738859.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Alm RA,  Smith D;
XX
DR   WPI: 1997-503122/46.
XX
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
XX
PS   Example; Page 121; 114Spp; English.
XX
CC   This sequence represents a primer for the H.pylori murI gene.
CC   The amplified sequence was used to compare homology of the coding
CC   sequences of the invention with other known proteins. The protein encoded
CC   by the DNA of the invention may be used in a vaccine to prevent or treat
CC   H.pylori infection or to identify H.pylori polypeptide binding compounds,
CC   useful as potential H.pylori life cycle activators or inhibitors. The DNA
CC   and probes derived from it may be used for the identification of H.pylori
CC   in a sample and the diagnosis of H.pylori infection. Nucleic acid
CC   sequences complementary to the DNA act as antisense sequences and can be
CC   used to prevent the translation of H.pylori mRNA. Antibodies against the
CC   protein can be used in immunoassays to evaluate the abundance and
CC   distribution of H.pylori-specific antigens. The genomic sequence of
CC   H.pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H.pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H.pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
XX
SQ   Sequence 29 BP; 11 A; 2 C; 6 G; 10 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 18;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5798 caaaaacgcctattttcat 5816
           |||||||||||||||||||
Db      29 CAAAAACGCCTATTTTCAT 11
```



```
RESULT  85
Q87899/c
ID   Q87899 standard; DNA; 35 BP.
XX
AC   Q87899;
XX
DT   29-NOV-1995  (first entry)
XX
DE   Normalised library first strand cDNA synthesis primer.
XX
KW   Normalised cDNA library; directionally cloned cDNA library;
KW   screening; hybridisation; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    15..18
FT                   /*tag=  a
FT                   /note= "characteristic sequence identifier"
XX
PN   WO9508647-A.
XX
PD   30-MAR-1995.
XX
PF   23-SEP-1994;   94WO-US10821.
XX
PR   24-SEP-1993;   93US-0126594.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Efstratiadis A,  Soares MB;
XX
DR   WPI; 1995-139615/18.
XX
PT   New normalised directional cDNA libraries - used for isolating
PT   novel cDNA's, including tissue-specific and development-specific
PT   DNA.
XX
PS   Disclosure; Page 45; 186pp; English.
XX
CC   Human tissues were obtained for construction of a variety of cDNA
CC   libraries, including infant brain, adult brain and adult
CC   hippocampus. Each of the cDNA libraries had a characteristic
CC   sequence identifier, provided by the oligonucleotide utilised to
CC   prime first strand cDNA synthesis (see Q87894-Q87907 for these
CC   primer sequences; all these primers have the PacI restriction site
CC   for directional cloning of cDNAs). Each of the libraries was
CC   propagated in the form of single-stranded (ss) circles and
CC   normalised separately by a novel method. The method comprises:
```



```
CC   generating fragments complementary to the 3' non-coding sequence of
CC   the ss circles in the library to produce partial duplexes;
CC   purifying the partial duplexes; melting and reassociating them to
CC   appropriate Cot; and purifying the unassociated ss circles to
CC   generate a normalised cDNA library.
XX
SQ   Sequence 35 BP; 6 A; 1 C; 1 G; 27 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 16;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgctt 407
           |||||||||||||||||||
Db      32 AAAAAAAAAAAAAATGCTT 14


RESULT  86
X97725
ID   X97725 standard; DNA; 625 BP.
XX
AC   X97725;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Extended human secreted protein coding sequence, SEQ ID NO. 290.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
```

```
DR   WPI; 1999-385906/32.
DR   P-PSDB; Y36041.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 364; 516pp; English.
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 625 BP; 138 A; 181 C; 114 G; 188 T; 4 other;

  Query Match             0.3%;  Score 19;  DB 20;  Length 625;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 401
          |||||||||||||||||||
Db    607 ttaatgaaaaaaaaaaaaa 625


RESULT  87
Z52996
ID   Z52996 standard; cDNA; 760 BP.
XX
AC   Z52996;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Human prostate tumor cDNA library derived EST fragment #139.
XX
KW   Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW   treatment; ds.
XX
OS   Homo sapiens.
XX
PN   DE19820190-A1.
XX
PD   04-NOV-1999.
```



```
XX
PF   28-APR-1998;   98DE-1020190.
XX
PR   28-APR-1998;   98DE-1020190.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-621386/54.
DR   P-PSDB; Y74232, Y74233, Y74234, Y74235, Y74236, Y74237.
XX
PT   New human nucleic acid sequences from pancreatic tumors, and related
PT   proteins
XX
PS   Claim 2; Page 293; 502pp; German.
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. Z52858-Z53014 represent
CC   expressed sequence tag (EST) fragments derived from a human pancreatic
CC   tumor cDNA library and which encode the proteins represented in
CC   Y73814-Y74252.
XX
SQ   Sequence 760 BP; 250 A; 126 C; 136 G; 248 T; 0 other;

  Query Match             0.3%;  Score 19;  DB 20;  Length 760;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    384 taatgaaaaaaaaaaaaa 402
          |||||||||||||||||||
Db    737 taatgaaaaaaaaaaaaa 755


RESULT  88
Z39022/c
ID   Z39022 standard; cDNA; 801 BP.
XX
AC   Z39022;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Mouse novel cDNA clone 18/25 3'-end Ese-SH3 interacting clone.
XX
KW   Mouse; murine; Esel; Ese2; endocytosis; vesicular trafficking;
KW   regulation; actin cytoskeleton; detection; cancer; infection;
KW   EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW   antiproliferative; antiviral; ss.
```



```
XX
OS   Mus sp.
XX
PN   WO9955728-A2.
XX
PD   04-NOV-1999.
XX
PF   27-APR-1999;   99WO-CA00375.
XX
PR   27-APR-1998;   98CA-2230201.
PR   05-FEB-1999;   99US-0118739.
XX
PA   (HSCR-) HSC RES & DEV LP.
XX
PI   Egan SE,  Wang W,  Sengar A;
XX
DR   WPI; 2000-052802/04.
XX
PT   New nucleic acid encoding Esel and 2 proteins, involved in regulation
PT   of endocytosis, used e.g. for treating cancer or preventing viral
PT   infection
XX
PS   Disclosure; Page 55-56; 99pp; English.
XX
CC   The present invention specifically describes mammalian Esel and 2
CC   proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
CC   regulator of endocytosis). (I) are involved in regulation of clathrin-
CC   mediated endocytosis (as a complex with Esp15 protein), vesicular
CC   trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
CC   mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
CC   sequences antisense to the (I) polynucleotide; agents that downregulate
CC   expression of Ese genes or antagonists of an Ese binding partner are
CC   used to treat diseases associated with undesirable endocytosis and
CC   resulting changes in cellular function. Particularly overexpression of
CC   Esel is used to block clathrin-mediated endocytosis in vivo or in cell
CC   cultures, while administration of (I) is used to promote endocytosis of
CC   selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
CC   proliferation of cells that can be stimulated to proliferate by a growth
CC   factor receptor; and similar compounds (also inactive Ese mutants) can be
CC   used to prevent viral infection. Endocytosis may also be regulated, in
CC   vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
CC   dynamin to the complex. Generally conditions that can be treated include
CC   cancer; abnormal cell division or migration; viral infection; or abnormal
CC   receptor signalling, tissue development or synaptic transmission. The
CC   present sequence represents a mouse cDNA clone Ese-SH3 interacting clone.
XX
SQ   Sequence 801 BP; 253 A; 123 C; 142 G; 283 T; 0 other;

  Query Match             0.3%;  Score 19;  DB 21;  Length 801;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
```



```
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    384 taatgaaaaaaaaaaaaa 402
          |||||||||||||||||||
Db     38 TAATGAAAAAAAAAAAAA 20


RESULT  89
V24653/c
ID   V24653 standard; DNA; 879 BP.
XX
AC   V24653;
XX
DT   26-JUN-1998  (first entry)
XX
DE   H. pylori ORF 05ae20220orf58.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   bacterium; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..879
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   WO9737044-A1.
XX
PD   09-OCT-1997.
XX
PF   27-MAR-1997;   97WO-US05223.
XX
PR   06-DEC-1996;   96US-0761318.
PR   29-MAR-1996;   96US-0625811.
PR   02-APR-1996;   96US-0758731.
PR   25-OCT-1996;   96US-0736905.
PR   28-OCT-1996;   96US-0738859.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Alm RA,  Smith D;
XX
DR   WPI; 1997-503122/46.
DR   P-PSDB; W55244.
XX
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
```

XX
PS Claims 5,6; Page 164; 1145pp; English.
XX
CC This sequence encodes a Helicobacter pylori protein of unspecified
CC function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The DNA and probes derived from it may be used for the
CC identification of H. pylori in a sample, and the diagnosis of
CC H. pylori infection. Nucleic acid sequences complementary to the
CC DNA act as antisense sequences, and can be used to prevent the
CC translation of H. pylori mRNA. Antibodies against the protein can
CC be used in immunoassays to evaluate the abundance and distribution
CC of H. pylori-specific antigens. The genomic sequence of H. pylori
CC (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were
CC analysed for ORF of at least 180 nucleotides, and the predicted
CC coding regions defined by computer evaluation. To identify likely
CC H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having
CC identified and determined the sequences of interest, particular
CC regions can be isolated from H. pylori by PCR amplification for
CC recombinant polypeptide production, e.g. in E. coli hosts.
XX
SQ Sequence 879 BP; 292 A; 130 C; 198 G; 259 T; 0 other;

Query Match 0.3%; Score 19; DB 18; Length 879;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5050 ttacaatttgagccattct 5068
||||||||||||||||||
Db 19 TTACAATTTGAGCCATTCT 1

RESULT 90
Z00843/c
ID Z00843 standard; cDNA; 991 BP.
XX
AC Z00843;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human secreted protein cDNA endoding gene 42.
XX
KW Secreted protein; prevention; treatment; protein therapy; gene therapy;
KW diagnosis; cancer; tumour; neurodegenerative disorder; blood disorder;
KW developmental abnormality; fetal deficiency; leukemia; autoimmune; acne;

KW hepatic disease; renal disease; lymphoma; inflammation; allergy; asthma;
KW Alzheimer's disease; cognitive disorder; schizophrenia; obesity; sepsis;
KW osteoporosis; arthritis; infection; AIDS; connective tissue disorder;
KW transplant rejection; diabetes; psoriasis; cardiovascular disorder;
KW reproductive disorder; food additive; food preservative; human; primer;
KW early promoter; GAS; gamma activation element; ss.
XX
OS Homo sapiens.
XX
PN WO9940100-A1.
XX
PD 12-AUG-1999.
XX
PF 04-FEB-1999; 99WO-US02293.
XX
PR 05-FEB-1998; 98US-0074341.
PR 09-FEB-1998; 98US-0074037.
PR 09-FEB-1998; 98US-0074118.
PR 09-FEB-1998; 98US-0074141.
PR 09-FEB-1998; 98US-0074157.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kyaw H, Lafleur DW, Moore PA, Rosen CA, Ruben SM;
PI Shi Y, Wei Y;
XX
DR WPI; 1999-479426/40.
DR P-PSDB; Y30852.
XX
PT New isolated human genes potentially useful for, e.g. developmental
PT abnormalities and fetal deficiencies
XX
PS Claim 1a; Page 199; 263pp; English.
XX
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The polynucleotides and their corresponding
CC secreted polypeptides are useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. Also pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for the
CC polynucleotides of the invention based on which tissues they are most
CC highly expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and fetal deficiencies, blood disorders, leukemias,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC disease, lymphomas, inflammation, allergies, Alzheimer's and cognitive
CC disorders, schizophrenia, obesity, osteoporosis, arthritis, infections,
CC AIDS, connective tissue disorders, transplant rejection, diabetes,
CC asthma, sepsis, acne, psoriasis, cardiovascular disorders, and
CC reproductive disorders. The polypeptides or polynucleotides can also be

CC used as food additives or preservatives. The polypeptide are also useful
CC for identifying their binding partners. This sequence encodes a
CC secreted protein described in the invention.
XX
SQ Sequence 991 BP; 269 A; 174 C; 174 G; 374 T; 0 other;

Query Match 0.3%; Score 19; DB 20; Length 991;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaat 403
||||||||||||||||||
Db 144 AATGAAAAAAAAAAAAAT 126

RESULT 91
X00713
ID X00713 standard; DNA; 1015 BP.
XX
AC X00713;
XX
DT 25-MAR-1999 (first entry)
XX
DE Human secreted protein gene 51 clone HAPNO80.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9842738-A1.
XX
PD 01-OCT-1998.
XX
PF 19-MAR-1998; 98WO-US05311.
XX
PR 30-MAY-1997; 97US-0050937.
PR 21-MAR-1997; 97US-0041276.
PR 21-MAR-1997; 97US-0041277.
PR 21-MAR-1997; 97US-0041281.
PR 21-MAR-1997; 97US-0042344.
PR 30-MAY-1997; 97US-0048069.
PR 30-MAY-1997; 97US-0048094.
PR 30-MAY-1997; 97US-0048095.

PR 30-MAY-1997; 97US-0048096.
PR 30-MAY-1997; 97US-0048099.
PR 30-MAY-1997; 97US-0048131.
PR 30-MAY-1997; 97US-0048135.
PR 30-MAY-1997; 97US-0048154.
PR 30-MAY-1997; 97US-0048160.
PR 30-MAY-1997; 97US-0048186.
PR 30-MAY-1997; 97US-0048187.
PR 30-MAY-1997; 97US-0048188.
PR 30-MAY-1997; 97US-0048350.
PR 30-MAY-1997; 97US-0048351.
PR 30-MAY-1997; 97US-0048352.
PR 30-MAY-1997; 97US-0048355.
PR 05-AUG-1997; 97US-0054804.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
PI Rosen CA, Ruben SM, Shi Y, Young P;
XX
DR WPI; 1999-070066/06.
DR P-PSDB; W67909.
XX
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 262; 385pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. X00602) for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 87 novel genes and their fragments (nucleic acid
CC sequences: X00611-X00724: amino acid sequences W67807-W68004) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 87 polynucleotides, based on
CC which tissues they are most highly expressed in (see X00611 for described
CC uses).
XX
SQ Sequence 1015 BP; 274 A; 248 C; 298 G; 194 T; 1 other;

Query Match 0.3%; Score 19; DB 20; Length 1015;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

```
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     387 tgaaaaaaaaaaaaaatgc 405
           ||||||||||||||||||
Db     868 tgaaaaaaaaaaaaaatgc 886


RESULT  92
V82785
ID   V82785 standard; cDNA; 1046 BP.
XX
AC   V82785;
XX
DT   25-FEB-1999  (first entry)
XX
DE   Clone ek591_1 isolated from human foetal brain cDNA library.
XX
KW   Secreted protein; nutritional activity; immune stimulating; vaccine;
KW   suppressing activity; haematopoiesis regulating activity;
KW   tissue growth activity; activin; inhibin activity; chemotactaxis;
KW   chemokinetic activity; haemostasis; thrombolytic activity; receptor;
KW   ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW   tumour inhibition; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO9842739-A2.
XX
PD   01-OCT-1998.
XX
PF   20-MAR-1998;   98WO-US05653.
XX
PR   19-MAR-1998;   98US-0044466.
PR   21-MAR-1997;   97US-0822167.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1998-609890/51.
DR   P-PSDB; W85463, W85464.
XX
PT   New polynucleotides encoding secreted human proteins - derived from
PT   human foetal brain, adult brain, foetal kidney, placenta or adult
PT   pineal gland cDNA libraries.
XX
PS   Claim 32; Page 83-84; 113pp; English.
XX
CC   The present sequence encodes a secreted protein (two potential
CC   reading frames have been identified). The polynucleotide and
```



```
CC   secreted protein are predicted to have biological activities which would
CC   make them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is given.
CC   Suggested activities include nutritional activity, immune stimulating
CC   (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC   invasion suppressor activity, and tumour inhibition activity (no data is
CC   given in the specification to support these activities). The
CC   polynucleotide is also stated to be useful for gene therapy.
XX
SQ   Sequence 1046 BP; 268 A; 221 C; 267 G; 289 T; 1 other;


  Query Match             0.3%;  Score 19;  DB 19;  Length 1046;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     4633 aaaaatacggcttcaagaa 4651
            ||||||||||||||||||
Db      325 aaaaatacggcttcaagaa 343


RESULT  93
V53421
ID   V53421 standard; DNA; 1213 BP.
XX
AC   V53421;
XX
DT   30-OCT-1998  (first entry)
XX
DE   DNA encoding a Staphylococcus aureus protein of unknown function.
XX
KW   Staphylococcus aureus protein; immune response induction; eye infection;
KW   antibody production; T-cell immune response; gastrointestinal infection;
KW   respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW   central nervous system; kidney infection; urinary tract infection;
KW   antimicrobial compound identification; broad spectrum antibiotic;
KW   therapy; ss.
XX
OS   Staphylococcus aureus.
XX
PN   EP841394-A2.
XX
PD   13-MAY-1998.
XX
PF   24-SEP-1997;   97EP-0307485.
XX
PR   24-SEP-1996;   96US-0027032.
XX
```



```
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT,  Burnham MKR,  Hodgson JE,  Knowles DJC;
PI   Lonetto MA,  Nicholas RO,  Pratt JM,  Reichard RW,  Rosenberg M;
PI   Ward JM;
XX
DR   WPI; 1998-252940/23.
DR   P-PSDB; W77626.
XX
PT   New nucleic acid sequences from Staphylococcus aureus WCHU29 -
PT   useful in vaccines and for treatment of bacterial infections of e.g.
PT   respiratory tract and central nervous system
XX
PS   Claim 1; Page 110; 390pp; English.
XX
CC   This sequence encodes a Staphylococcus aureus protein of unknown
CC   function, and represents a DNA Sequence of the invention.
CC   The DNA sequences were isolated from Staphylococcus aureus WCHU29
CC   (NCIMB 40771). Host cells containing the DNA sequences are used to
CC   produce polypeptides or fragments. The proteins are used in the treatment
CC   of disease, for inducing an immune response by administering them, to
CC   produce antibody and/or T-cell immune response. Antagonists of the
CC   proteins are used for the inhibition of bacterial polypeptides.
CC   Conditions which may be treated include bacterial infections, especially
CC   respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
CC   urinary tract, skin, bones and joints. The proteins can also be used to
CC   identify antimicrobial compounds which are broad spectrum antibiotics,
CC   especially useful in the treatment of H. pylori infection.
XX
SQ   Sequence 1213 BP; 460 A; 175 C; 258 G; 320 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 19;  Length 1213;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     373 ttaataatgattaatgaaa 391
           ||||||||||||||||||
Db     261 ttaataatgattaatgaaa 279


RESULT  94
X00661
ID   X00661 standard; DNA; 1580 BP.
XX
AC   X00661;
XX
DT   25-MAR-1999  (first entry)
XX
DE   Human secreted protein gene 51 clone HAPNO80.
```



```
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9842738-A1.
XX
PD   01-OCT-1998.
XX
PF   19-MAR-1998;   98WO-US05311.
XX
PR   30-MAY-1997;   97US-0050937.
PR   21-MAR-1997;   97US-0041276.
PR   21-MAR-1997;   97US-0041277.
PR   21-MAR-1997;   97US-0041281.
PR   21-MAR-1997;   97US-0042344.
PR   30-MAY-1997;   97US-0048069.
PR   30-MAY-1997;   97US-0048094.
PR   30-MAY-1997;   97US-0048095.
PR   30-MAY-1997;   97US-0048096.
PR   30-MAY-1997;   97US-0048099.
PR   30-MAY-1997;   97US-0048131.
PR   30-MAY-1997;   97US-0048135.
PR   30-MAY-1997;   97US-0048154.
PR   30-MAY-1997;   97US-0048160.
PR   30-MAY-1997;   97US-0048186.
PR   30-MAY-1997;   97US-0048187.
PR   30-MAY-1997;   97US-0048188.
PR   30-MAY-1997;   97US-0048350.
PR   30-MAY-1997;   97US-0048351.
PR   30-MAY-1997;   97US-0048352.
PR   30-MAY-1997;   97US-0048355.
PR   05-AUG-1997;   97US-0054804.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Duan R,  Ebner R,  Ferrie AM,  Florence KA;
PI   Greene JM,  Hu JS,  Lafleur DW,  Moore PA,  Ni J,  Olsen HS;
PI   Rosen CA,  Ruben SM,  Shi Y,  Young P;
XX
DR   WPI; 1999-070066/06.
DR   P-PSDB; W67857.
XX
PT   New isolated human genes and the secreted polypeptides they encode -
```

```
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 213-214; 385pp; English.
XX
CC   This sequence represents a nucleic acid molecule which encodes a secreted
CC   human protein.  The gene number, and the clone it is derived from, are
CC   detailed in the descriptor line.  The gene can be used to generate fusion
CC   proteins by linking to the gene to a human immunoglobulin Fc portion
CC   (e.g. X00602) for increasing the stability of the fused protein as
CC   compared to the human protein only.
CC   The invention relates to 87 novel genes and their fragments (nucleic acid
CC   sequences: X00611-X00724; amino acid sequences W67807-W68004) which
CC   are useful for preventing, treating or ameliorating  medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 87 polynucleotides, based on
CC   which tissues they are most highly expressed in (see X00611 for described
CC   uses).
XX
SQ   Sequence 1580 BP; 357 A; 448 C; 479 G; 294 T; 2 other;


  Query Match             0.3%;  Score 19;  DB 20;  Length 1580;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     387 tgaaaaaaaaaaaaatgc 405
           ||||||||||||||||||
Db    1440 tgaaaaaaaaaaaaatgc 1458


RESULT  95
X14271/c
ID   X14271 standard; DNA; 1786 BP.
XX
AC   X14271;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 1629 gene.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease; ss.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..1736
FT                   /*tag= a
```



```
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   P-PSDB; W98552.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 1; Page 1073-1076; 2054pp; English.
XX
CC   This sequence represents a polynucleotide of the invention. It was
CC   isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC   The polypeptides can be used for preventing or treating Helicobacter
CC   infections, and gastroduodenal diseases associated with these
CC   infections, including acute, chronic, and atrophic gastritis, and peptic
CC   ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC   for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence 1786 BP; 557 A; 325 C; 394 G; 510 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 19;  Length 1786;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    4018 ctaaaagcgaggggttttt 4036
           ||||||||||||||||||||
Db     156 CTAAAAGCGAGGGGTTTTT 138


RESULT  96
X00621
ID   X00621 standard; DNA; 2042 BP.
XX
AC   X00621;
```



```
XX
DT   25-MAR-1999  (first entry)
XX
DE   Human secreted protein gene 11 clone HCENE16.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9842738-A1.
XX
PD   01-OCT-1998.
XX
PF   19-MAR-1998;   98WO-US05311.
XX
PR   30-MAY-1997;   97US-0050937.
PR   21-MAR-1997;   97US-0041276.
PR   21-MAR-1997;   97US-0041277.
PR   21-MAR-1997;   97US-0041281.
PR   21-MAR-1997;   97US-0042344.
PR   30-MAY-1997;   97US-0048069.
PR   30-MAY-1997;   97US-0048094.
PR   30-MAY-1997;   97US-0048095.
PR   30-MAY-1997;   97US-0048096.
PR   30-MAY-1997;   97US-0048099.
PR   30-MAY-1997;   97US-0048131.
PR   30-MAY-1997;   97US-0048135.
PR   30-MAY-1997;   97US-0048154.
PR   30-MAY-1997;   97US-0048160.
PR   30-MAY-1997;   97US-0048186.
PR   30-MAY-1997;   97US-0048187.
PR   30-MAY-1997;   97US-0048188.
PR   30-MAY-1997;   97US-0048350.
PR   30-MAY-1997;   97US-0048351.
PR   30-MAY-1997;   97US-0048352.
PR   30-MAY-1997;   97US-0048355.
PR   05-AUG-1997;   97US-0054804.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Duan R,  Ebner R,  Ferrie AM,  Florence KA;
PI   Greene JM,  Hu JS,  Lafleur DW,  Moore PA,  Ni J,  Olsen HS;
PI   Rosen CA,  Ruben SM,  Shi Y,  Young P;
XX
```



```
DR   WPI; 1999-070066/06.
DR   P-PSDB; W67817.
XX
PT   New isolated human genes and the secreted polypeptides they encode -
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 177-178; 385pp; English.
XX
CC   This sequence represents a nucleic acid molecule which encodes a secreted
CC   human protein.  The gene number, and the clone it is derived from, are
CC   detailed in the descriptor line.  The gene can be used to generate fusion
CC   proteins by linking to the gene to a human immunoglobulin Fc portion
CC   (e.g. X00602) for increasing the stability of the fused protein as
CC   compared to the human protein only.
CC   The invention relates to 87 novel genes and their fragments (nucleic acid
CC   sequences: X00611-X00724; amino acid sequences W67807-W68004) which
CC   are useful for preventing, treating or ameliorating  medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 87 polynucleotides, based on
CC   which tissues they are most highly expressed in (see X00611 for described
CC   uses).
XX
SQ   Sequence 2042 BP; 676 A; 372 C; 411 G; 582 T; 1 other;


  Query Match             0.3%;  Score 19;  DB 20;  Length 2042;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     384 taatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2013 taatgaaaaaaaaaaaaaa 2031


RESULT  97
A26310
ID   A26310 standard; cDNA; 2457 BP.
XX
AC   A26310;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 30 SEQ ID NO:40.
XX
KW   Human; secreted protein; diagnosis; neuroprotective; nootropic;
KW   neuroleptic; antimanic; cerebroprotective; immunomodulatory;
KW   anti-microbial; cardiant; cytostatic; antiinflammatory; haemostatic;
KW   anticonvulsant; vasotropic; vaccine; gene therapy; anti-sense therapy;
```

```
KW   neural; reproductive; immune disorder; immunodeficiency; infection;
KW   lymphoma; demyelinating disease; autoimmunity; cancer; inflammation;
KW   aneurysm; haemorrhage; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; Tourette syndrome; multiple sclerosis; meningitis;
KW   ischaemia; mania; dementia; obsessive compulsive disorder;
KW   viral prophylaxis; developmental disorder; sexually-linked disorder;
KW   cardiovascular disorder; food additive; preservative; chromosome 11; ss.
XX
OS   Homo sapiens.
XX
PN   WO200011014-A1.
XX
PD   02-MAR-2000.
XX
PF   24-AUG-1999;   99WO-US19330.
XX
PR   25-AUG-1998;   98US-0097917.
PR   31-AUG-1998;   98US-0098634.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Moore PA,  Ruben SM,  Olsen HS,  Shi Y,  Rosen CA,  Florence KA;
PI   Soppet DR,  Lafleur DW,  Endress GA,  Ebner R,  Komatsoulis G;
PI   Duan RD;
XX
DR   WPI; 2000-224656/19.
DR   P-PSDB; Y91375.
XX
PT   Novel secreted proteins and corresponding DNA molecules that can be
PT   used to prevent, treat and diagnose disease in humans, for example,
PT   Alzheimer's, cancer, and immune disorders -
XX
PS   Claim 1; Page 332-333; 416pp; English.
XX
CC   The polynucleotide sequences given in A26281 to A26336 encode the human
CC   secreted proteins given in Y91346 to Y91449. The human secreted proteins
CC   can have activities based on the tissues and cells they are expressed in.
CC   Examples of the activities are: neuroprotective; nootropic; neuroleptic;
CC   antimanic; cerebroprotective; immunomodulatory; anti-microbial; cardiant;
CC   cytostatic; antiinflammatory; haemostatic; anticonvulsant; and
CC   vasotropic. The polynucleotides and proteins may be used to prevent,
CC   treat or ameliorate a medical condition, e.g. by protein or gene therapy.
CC   Conditions treatable by the proteins of the invention include neural,
CC   reproductive, or immune disorders, especially immunodeficiency,
CC   infection, lymphomas, demyelinating diseases, auto-immunities, cancer,
CC   general microbial infection, inflammation, aneurysms and haemorrhages.
CC   Specific examples include: Alzheimer's disease; Parkinson's;
CC   Huntington's; Tourette syndrome; multiple sclerosis; meningitis;
CC   ischaemia; prostate cancer; mania; dementia; obsessive compulsive
CC   disorder and viral prophylaxis. The polynucleotides and proteins can also
CC   be used in the detection of disorders associated with the function of the
```



```
CC   protein, for example, the detection of developmental disorders,
CC   sexually-linked disorders, or disorders of the cardiovascular system.
CC   They may also be used as food additives or preservatives. A26272 to
CC   A26280 and Y91345 are sequences used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 2457 BP; 710 A; 471 C; 527 G; 745 T; 4 other;


  Query Match           0.3%;  Score 19;  DB 21;  Length 2457;
  Best Local Similarity  100.0%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     384 taatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2396 taatgaaaaaaaaaaaaaa 2414


RESULT  98
V61484/c
ID   V61484 standard; cDNA; 2547 BP.
XX
AC   V61484;
XX
DT   11-JAN-1999  (first entry)
XX
DE   Human secreted protein dn904_2 cDNA.
XX
KW   Secreted protein; human; dn904_2; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1563..1688
FT                   /*tag= a
XX
PN   WO9841539-A2.
XX
PD   24-SEP-1998.
XX
PF   19-MAR-1998;   98WO-US05474.
XX
PR   18-MAR-1998;   98US-0040963.
PR   19-MAR-1997;   97US-0820493.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  LaVallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1998-521163/44.
```



```
DR   P-PSDB; W79093.
XX
PT   New polynucleotide(s) encoding secreted human proteins - derived
PT   from human foetal kidney, adult testes and adult or foetal brain
PT   cDNA libraries
XX
PS   Claim 27; Page 80-82; 112pp; English.
XX
CC   This full-length cDNA clone, designated dn904_2, codes for a novel
CC   secreted human protein (see W79093).  It was isolated from a human
CC   foetal brain cDNA library using methods which are selective for
CC   cDNAs encoding secreted proteins, or was identified as encoding
CC   a secreted or transmembrane protein on the basis of computer
CC   analysis of the encoding protein.  The nucleotide sequence shows
CC   homology to some database sequences, and may contain an Alu
CC   repetitive element.  The invention provides cDNA clones (see
CC   V61477-87) from human foetal kidney, adult testis, and adult or
CC   foetal brain cDNA libraries that code for secreted proteins (see
CC   W79087-97).  These clones are deposited as ATCC 98364.  The
CC   polynucleotides and proteins are predicted to have useful
CC   biological activities which would make them suitable for treating,
CC   preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given.  Suggested
CC   activities include nutritional, immune stimulating (e.g. as
CC   vaccines) or suppressing, haematopoiesis regulating, tissue growth,
CC   activin/inhibin, chemotactic/chemokinetic, haemostatic and
CC   thrombolytic, receptor/ligand, antiinflammatory, cadherin/tumour
CC   invasion suppressor and tumour inhibition activities.  The
CC   polynucleotides are also stated to be useful for gene therapy.
XX
SQ   Sequence 2547 BP; 640 A; 476 C; 416 G; 1012 T; 3 other;


  Query Match           0.3%;  Score 19;  DB 19;  Length 2547;
  Best Local Similarity  100.0%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3020 atgctagaaaaaaatctga 3038
           ||||||||||||||||||||
Db     333 ATGCTAGAAAAAATCTGA 315


RESULT  99
T58898/c
ID   T58898 standard; cDNA; 2883 BP.
XX
AC   T58898;
XX
DT   28-APR-1997  (first entry)
XX
DE   C-Delta-1 gene (alternatively spliced variant).
```



```
XX
KW   C-Delta-1; cell proliferation; nervous system disorder;
KW   tissue regeneration; Notch; cervix cancer; breast cancer;
KW   lung cancer; colon cancer; melanoma; seminoma;
KW   neurogenesis; therapy; ss.
XX
OS   Gallus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             277..2499
FT                   /*tag= a
XX
PN   WO9701571-A1.
XX
PD   16-JAN-1997.
XX
PF   28-JUN-1996;   96WO-US11178.
XX
PR   28-JUN-1995;   95US-0000589.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PA   (UYYA ) UNIV YALE.
XX
PI   Artavanis-Tsakonas S,  Gray GE,  Henrique D,  Ish-Horowicz D;
PI   Lewis J;
XX
DR   WPI; 1997-100159/09.
DR   P-PSDB; W00876.
XX
PT   New vertebrate Delta protein, DNA and antibodies - for treating and
PT   preventing cancer, nervous system disorders and for tissue
PT   regeneration
XX
PS   Disclosure; Fig 1B1-1B2; 135pp; English.
XX
CC   The C-delta-1 gene (T58898) codes for the chick homologue (W00876)
CC   of Drosophila Delta, a protein that binds to Notch protein.  It was
CC   obtd. by PCR amplification of cDNA from stage 4-6 embryos using
CC   primers (see also T59455-56) based on fly Delta sequences.  A
CC   shorter variant (T58897) was also identified.  C-Delta-1 expression
CC   appears to be the earliest known marker for prospective neurons.
CC   Mouse (T58899) and human (T58900, T59454) Delta-1 sequences have
CC   also been isolated.  Delta-1 genes can be used in the prodn. of
CC   Delta polypeptides and (including antisense sequences) utilised in
CC   the treatment of disorders of cell fate or differentiation, such as
CC   cancer, and nervous system disorders, or to promote tissue
CC   regeneration and repair.
XX
SQ   Sequence 2883 BP; 638 A; 858 C; 792 G; 593 T; 2 other;
```

```
  Query Match             0.3%;  Score 19;  DB 18;  Length 2883;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3220 gttaataaaaagaaagcag 3238
           ||||||||||||||||||||
Db    2818 GTTAATAAAAAGAAAGCAG 2800


RESULT 100
T16857
ID   T16857 standard; cDNA; 4089 BP.
XX
AC   T16857;
XX
DT   20-MAY-1996  (first entry)
XX
DE   Bovine endothelial nitric oxide synthase cDNA clone pBOS13.
XX
KW   Nitric oxide synthase; NOS; endothelial cell; diagnosis;
KW   gene therapy; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             30..3647
FT                   /*tag= a
FT                   /transl_except= pos:2187..2189:_aa:Ala
FT                   /transl_except= pos:2190..2192:_aa:Arg
XX
PN   US5498539-A.
XX
PD   12-MAR-1996.
XX
PF   02-JUL-1992;   92US-0908245.
XX
PR   02-JUL-1992;   92US-0908245.
XX
PA   (DAUC ) DAIICHI PHARM CO LTD.
XX
PI   Alexander RW,  Harrison DG,  Murphy TJ,  Nishida K;
XX
DR   WPI; 1996-159689/16.
DR   P-PSDB; R93930.
XX
PT   DNA encoding bovine endothelial nitric oxide synthase - used as
PT   probe for research and diagnostic purposes, and therapeutically to
PT   increase nitric oxide prodn. in blood vessels
XX
PS   Claim 2; Column 15-20; 21pp; English.
XX
```



```
CC   A cDNA clone (T16857), designated pBOS13, codes for bovine
CC   endothelial nitric oxide synthase (NOS) (R93930).  pBOS13 was
CC   isolated from a cDNA library prepd. from bovine aortic endothelial
CC   cells by screening with a fragment of rat brain NOS cDNA.  The NOS
CC   gene can be used in the prepn. of probes (see T16858-60), or may be
CC   transfected into host cells for prodn. of large vols. of endothelial
CC   NOS.  Alternatively, the gene is transfected into blood vessels in
CC   vivo to give a therapeutic increase in nitric oxide prodn.
XX
SQ   Sequence 4089 BP; 761 A; 1400 C; 1237 G; 691 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 17;  Length 4089;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     384 taatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    4071 taatgaaaaaaaaaaaaaa 4089


RESULT 101
X76320/c
ID   X76320 standard; DNA; 4748 BP.
XX
AC   X76320;
XX
DT   05-AUG-1999  (first entry)
XX
DE   Schizosaccharomyces pombe invertase precursor inv1+ encoding DNA.
XX
KW   Schizosaccharomyces pombe; invertase precursor; mitotic yeast; inv1+;
KW   molecular entity; ds.
XX
OS   Schizosaccharomyces pombe.
XX
FH   Key             Location/Qualifiers
FT   CDS             2810..4555
FT                   /*tag= a
XX
PN   JP11127865-A.
XX
PD   18-MAY-1999.
XX
PF   31-OCT-1997;   97JP-0314607.
XX
PR   31-OCT-1997;   97JP-0314607.
XX
PA   (ASAG ) ASAHI GLASS CO LTD.
XX
DR   WPI; 1999-350326/30.
```



```
DR   P-PSDB; Y17500.
XX
PT   New invertase precursor - cloned from Schizosaccharomyces pombe
XX
PS   Claim 1; Page 6-9; 15pp; Japanese.
XX
CC   The present sequence encodes an invertase precursor (I) of a mitotic
CC   yeast, Schizosaccharomyces pombe, designated inv1+. The molecular entity
CC   of the invertase of the mitotic yeast, Schizosaccharomyces pombe was
CC   clarified.
XX
SQ   Sequence 4748 BP; 1474 A; 908 C; 820 G; 1546 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 20;  Length 4748;
  Best Local Similarity   100.0%;  Pred. No. 96;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1060 atccgaacggatcaaaagt 1078
           ||||||||||||||||||||
Db    501 ATCCGAACGGATCAAAAGT 483


RESULT 102
X57223/c
ID   X57223 standard; DNA; 4748 BP.
XX
AC   X57223;
XX
DT   26-JUL-1999  (first entry)
XX
DE   S. pombe invertase DNA.
XX
KW   Inducible promoter; invertase; catabolite inhibition; vector;
KW   extracellular expression; foreign gene; regulation expression; ss.
XX
OS   Schizosaccharomyces pombe.
XX
PN   WO9923223-A1.
XX
PD   14-MAY-1999.
XX
PF   30-OCT-1998;   98WO-JP04929.
XX
PR   31-OCT-1997;   97JP-0314608.
XX
PA   (ASAG ) ASAHI GLASS CO LTD.
XX
PI   Hama Y,  Kumagai H,  Tohda H;
XX
DR   WPI; 1999-313344/26.
```



```
DR   P-PSDB; Y08399.
XX
PT   Promoter and secretory signal sequences from invertase gene of
PT   Schizosaccharomyces pombe
XX
PS   Claim.2; Page 48-52; 67pp; Japanese.
XX
CC   This invention describes a novel promoter region of the moiety relating
CC   to catabolite inhibition of the invertase gene of Schizosaccharomyces
CC   pombe. The products of the invention can be used in the construction of
CC   a vector which allows extracellular expression of a foreign gene,
CC   regulated by the presence or absence of a specific nutrient such as
CC   glucose and results in the easy regulated expression of foreign genes
CC   in a yeast system.
XX
SQ   Sequence 4748 BP; 1474 A; 908 C; 820 G; 1546 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 20;  Length 4748;
  Best Local Similarity   100.0%;  Pred. No. 96;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1060 atccgaacggatcaaaagt 1078
           ||||||||||||||||||||
Db    501 ATCCGAACGGATCAAAAGT 483


RESULT 103
X24464/c
ID   X24464 standard; DNA; 5790 BP.
XX
AC   X24464;
XX
DT   07-JUN-1999  (first entry)
XX
DE   Bovine myostatin genomic DNA.
XX
KW   Myostatin; cattle; bovine; transforming growth factor beta;
KW   double muscling; muscle hyperplasia; transgenic animal; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             454..5439
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            454..826
FT                   /*tag= b
FT                   /number= 1
FT   intron          827..2652
FT                   /*tag= c
```

```
FT                   /number= 1
FT   exon            2653..3026
FT                   /*tag=  d
FT                   /number= 2
FT   intron          3027..5058
FT                   /*tag=  e
FT                   /number= 2
FT   exon            5059..5439
FT                   /*tag=  f
FT                   /number= 3
XX
PN   WO9902667-A1.
XX
PD   21-JAN-1999.
XX
PF   14-JUL-1998;    98WO-IB01197.
XX
PR   15-JAN-1998;    98US-0007761.
PR   14-JUL-1997;    97US-0891789.
XX
PA   (UYLI-) UNIV LIEGE.
XX
PI   Georges M,  Grobet L,  Poncelet D;
XX
DR   WPI; 1999-120869/10.
DR   P-PSDB; W97884.
XX
PT   Increasing muscle mass in mammals  - by decreasing myostatin
PT   expression
XX
PS   Claim 71; Page 66-70; 75pp; English.
XX
CC   This is the nucleotide sequence of genomic DNA encoding bovine
CC   myostatin (see W97884), a member of the transforming growth factor
CC   beta superfamily.  The sequence was determined following PCR
CC   amplification of genomic DNA using primers (see X24467-83) based
CC   on intron-exon boundaries predicted by comparison of murine and
CC   cattle gene sequences.  A mutant gene (see X24416) in which the
CC   coding sequence lacks an 11 bp consecutive sequence (see X24421)
CC   has also been sequenced.  Cattle of the Belgian Blue breed
CC   homozygous for the mutant gene are double-muscled.  A new method of
CC   increasing muscle mass of a mammal having myostatin-expressing
CC   muscle cells, comprises administration of a nucleic acid molecule
CC   substantially complementary to at least a portion of mRNA encoding
CC   myostatin and of sufficient length to reduce myostatin expression
CC   and thus increase muscle mass.  A ribozyme may also be used.  Also
CC   claimed are: a method for determining muscular hyperplasia (MH) in
CC   a mammal using primers based upstream and downstream of the
CC   mutation; a diagnostic kit for determining the genotype of a sample
CC   of genetic material; a method for determining MH in a mammal; a
CC   method for determining double muscling in a bovine animal; a method
```



```
CC   for determining the myostatin genotype of an animal; purified
CC   myostatin; isolated nucleic acids; a microbial host cell; a probe
CC   based on the myostatin gene mutation; transgenic mammals having MH
CC   phenotype; a myostatin knockout animal; and a transgenic bovine
CC   having a gene encoding active myostatin, and a heterologous
CC   nucleotide sequence antisense to that gene, and optionally further
CC   containing a gene encoding a nucleic acid sequence with ribozyme
CC   activity in transcriptional association with the antisense
CC   sequence.  Primers are preferably based on genomic bovine myostatin
CC   DNA and human myostatin cDNA (see X24418).
XX
SQ   Sequence 5790 BP; 1891 A; 1018 C; 1025 G; 1856 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 20;  Length 5790;
  Best Local Similarity   100.0%;  Pred. No. 95;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaa 400
           ||||||||||||||||||||
Db    1247 ATTAATGAAAAAAAAAAAA 1229


RESULT 104
Z21006
ID   Z21006 standard; DNA; 7179 BP.
XX
AC   Z21006;
XX
DT   30-NOV-1999  (first entry)
XX
DE   Salmonella dublin pathogenicity island SPI-5 DNA.
XX
KW   Pathogen; attenuation; food poisoning; enteritis; vaccine;
KW   inflammatory response; secretory response; ds.
XX
OS   Salmonella dublin.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..102
FT                   /*tag=  a
FT                   /partial
FT                   /product= "CopS protein"
FT   CDS             275..1576
FT                   /*tag=  b
FT                   /product= "pipD gene product"
FT   CDS             complement (1963..2133)
FT                   /*tag=  c
FT                   /product= "orfX gene product"
FT   CDS             2465..4150
FT                   /*tag=  d
```



```
FT                   /product= "SopB protein"
FT   CDS             4167..4508
FT                   /*tag=  e
FT                   /product= "pipC gene product"
FT   CDS             5057..5932
FT                   /*tag=  f
FT                   /product= "pipB gene product"
FT   CDS             6154..6834
FT                   /*tag=  g
FT                   /product= "pipA gene product"
XX
PN   WO9945120-A2.
XX
PD   10-SEP-1999.
XX
PF   08-MAR-1999;    99WO-GB00675.
XX
PR   06-MAR-1998;    98GB-0004809.
XX
PA   (BIOT-) BIOTECHNOLOGY & BIOLOGICAL SCI RES COUNC.
XX
PI   Wallis TS,  Galyov EE,  Wood MW;
XX
DR   WPI; 1999-540848/45.
DR   P-PSDB; Y42276, Y42277, Y42278, Y42279, Y42280, Y42281, Y42282.
XX
PT   A new attenuated Salmonella useful as a vaccine against Salmonella,
PT   Shigella or E. coli  -
XX
PS   Claim 1; Fig 3; 99pp; English.
XX
CC   This sequence represents a novel Salmonella dublin pathogenicity
CC   island, SPI-5. Pathogenicity islands are horizontally acquired clusters
CC   of genes, often inserted at tRNA loci, which confer certain virulence
CC   features on the host bacterium. SPI-5 encoded genes are necessary for the
CC   enteropathogenic phenotype of S. dublin. SPI-5 is conserved in Salmonella
CC   species, but was not found in other enteric bacteria tested. It is
CC   inserted between the tRNA-encoding serT gene and the copR/copS genes.
CC   SPI-5 encodes 6 genes. The SopB protein (Y42279) affects cellular
CC   responses, which leads to the influx of neutrophils into the intestinal
CC   epithelium and the induction of fluid secretion. Based on sequence
CC   similarity with other bacterial proteins, the putative pipD gene product
CC   (Y42277) may be a peptidase, having its target on the surface or inside
CC   of a host cell, while the putative pipB gene product (Y42281) may be
CC   involved in glycolipid biogenesis. The pipA gene product is also thought
CC   to be involved in pathogenicity. The introduction of mutations into one
CC   or more SPI-5 genes, particularly sopB, or into other
CC   enteropathogenocity-associated genes such as sopA (Z21007), can produce
CC   attenuated strains of enteric bacteria, e.g., Salmonella species. The
CC   enteropathogenicity of such attenuated bacteria is reduced in comparison
CC   to the wild-type, but the invasiveness is not substantially affected.
```



```
CC   Reduction in enteropathogenicity may be determined by measuring the
CC   intestinal secretory response and/or inflammatory response. Attenuated
CC   Salmonella bacteria can be used as a vaccine to treat or immunise against
CC   Salmonella, Shigella or E. coli infection, particularly in a person or
CC   agricultural animal, and to detect one of those organisms in a sample.
CC   Species of Salmonella include S. typhimurium, S. enteritidis, S. dublin
CC   and S. choleraesuis. The SPI-5 DNA and the proteins it encodes are used
CC   as immunogens, in the production of monoclonal antibodies and in the
CC   detection of the above Salmonella species. They can also be used to
CC   screen for agents which affect or modulate enteropathogenic protein
CC   activity.
XX
SQ   Sequence 7179 BP; 2190 A; 1603 C; 1547 G; 1839 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 20;  Length 7179;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2872 gtaacggataaagttgata 2890
           ||||||||||||||||||||
Db    5405 gtaacggataaagttgata 5423


RESULT 105
X57224/c
ID   X57224 standard; DNA; 7286 BP.
XX
AC   X57224;
XX
DT   26-JUL-1999  (first entry)
XX
DE   WO 9923223 Seq ID 2.
XX
KW   Inducible promoter; invertase; catabolite inhibition; vector;
KW   extracellular expression; foreign gene; regulation expression; ss.
XX
OS   Schizosaccharomyces pombe.
XX
PN   WO9923223-A1.
XX
PD   14-MAY-1999.
XX
PF   30-OCT-1998;    98WO-JP04929.
XX
PR   31-OCT-1997;    97JP-0314608.
XX
PA   (ASAG ) ASAHI GLASS CO LTD.
XX
PI   Hama Y,  Kumagai H,  Tohda H;
XX
```

```
DR   WPI; 1999-313344/26.
XX
PT   Promoter and secretory signal sequences from invertase gene of
PT   Schizosaccharomyces pombe
XX
PS   Disclosure; Page 52-57; 67pp; Japanese.
XX
CC   This invention describes a novel promoter region of the moiety relating
CC   to catabolite inhibition of the invertase gene of Schizosaccharomyces
CC   pombe. The products of the invention can be used in the construction of
CC   a vector which allows extracellular expression of a foreign gene,
CC   regulated by the presence or absence of a specific nutrient such as
CC   glucose and results in the easy regulated expression of foreign genes
CC   in a yeast system.
XX
SQ   Sequence 7286 BP; 2181 A; 1557 C; 1544 G; 2004 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 20;  Length 7286;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1060 atccgaacggatcaaaagt 1078
           ||||||||||||||||||||
Db   2016 ATCCGAACGGATCAAAAGT 1998


RESULT 106
X57235/c
ID   X57235 standard; DNA; 7938 BP.
XX
AC   X57235;
XX
DT   26-JUL-1999  (first entry)
XX
DE   WO 9923223 Seq ID 13.
XX
KW   Inducible promoter; invertase; catabolite inhibition; vector;
KW   extracellular expression; foreign gene; regulation expression; ss.
XX
OS   Schizosaccharomyces pombe.
XX
PN   WO9923223-A1.
XX
PD   14-MAY-1999.
XX
PF   30-OCT-1998;   98WO-JP04929.
XX
PR   31-OCT-1997;   97JP-0314608.
XX
PA   (ASAG ) ASAHI GLASS CO LTD.
```



```
XX
PI   Hama Y,  Kumagai H,  Tohda H;
XX
DR   WPI; 1999-313344/26.
XX
PT   Promoter and secretory signal sequences from invertase gene of
PT   Schizosaccharomyces pombe
XX
PS   Disclosure; Page 58-63; 67pp; Japanese.
XX
CC   This invention describes a novel promoter region of the moiety relating
CC   to catabolite inhibition of the invertase gene of Schizosaccharomyces
CC   pombe. The products of the invention can be used in the construction of
CC   a vector which allows extracellular expression of a foreign gene,
CC   regulated by the presence or absence of a specific nutrient such as
CC   glucose and results in the easy regulated expression of foreign genes
CC   in a yeast system.
XX
SQ   Sequence 7938 BP; 2338 A; 1781 C; 1726 G; 2093 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 20;  Length 7938;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1060 atccgaacggatcaaaagt 1078
           ||||||||||||||||||||
Db   2016 ATCCGAACGGATCAAAAGT 1998


RESULT 107
A09020/c
ID   A09020 standard; DNA; 14000 BP.
XX
AC   A09020;
XX
DT   01-AUG-2000  (first entry)
XX
DE   A. thaliana ATPAC genomic DNA.
XX
KW   Anion channel; ATPAC; mdr homologue; ATP binding cassette transporter;
KW   ABC transporter; p-glycoprotein; inducible; NPPB; herbicide resistance;
KW   transgenic plant; ds.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..3429
FT                   /*tag= a
FT                   /note= "promoter regulatory region"
FT   CDS             3429..10428
```



```
FT                   /*tag= b
FT   exon            3429..3683
FT                   /*tag= c
FT                   /number= 1
FT   intron          3684..3764
FT                   /*tag= d
FT                   /number= 1
FT   exon            3765..3819
FT                   /*tag= e
FT                   /number= 2
FT   intron          3820..4694
FT                   /*tag= f
FT                   /number= 2
FT   exon            4695..4870
FT                   /*tag= g
FT                   /number= 3
FT   intron          4871..5488
FT                   /*tag= h
FT                   /number= 3
FT   exon            5489..5689
FT                   /*tag= i
FT                   /number= 4
FT   intron          5690..5922
FT                   /*tag= j
FT                   /number= 4
FT   exon            5923..6182
FT                   /*tag= k
FT                   /number= 5
FT   intron          6183..6276
FT                   /*tag= l
FT                   /number= 5
FT   exon            6277..6593
FT                   /*tag= m
FT                   /number= 6
FT   intron          6594..6679
FT                   /*tag= n
FT                   /number= 6
FT   exon            6680..6903
FT                   /*tag= o
FT                   /number= 7
FT   intron          6904..6985
FT                   /*tag= p
FT                   /number= 7
FT   exon            6986..7805
FT                   /*tag= q
FT                   /number= 8
FT   intron          7806..7884
FT                   /*tag= r
FT                   /number= 8
FT   exon            7885..8126
FT                   /*tag= s
```



```
FT                   /number= 9
FT   intron          8127..9219
FT                   /*tag= t
FT                   /number= 9
FT   exon            9220..10428
FT                   /*tag= u
FT                   /number= 10
FT                   /transl_except= (pos:9388..9390, aa: Leu-Ser-Leu)
FT                   /transl_except= (pos:9859..9861, aa: Leu)
XX
PN   WO200018886-A1.
XX
PD   06-APR-2000.
XX
PF   24-SEP-1999;   99WO-US22363.
XX
PR   25-SEP-1998;   98US-0101814.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Spalding EP,  Noh B;
XX
DR   WPI; 2000-293135/25.
DR   P-PSDB; Y92173.
XX
PT   New Arabidopsis thaliana plPAC gene encoding an ABC transporter induced
PT   by exposure to NPPB, useful for detoxification of xenobiotics
XX
PS   Claim 6; Page 70-74; 78pp; English.
XX
CC   This is the Arabidopsis thaliana putative anion channel (ATPAC) genomic
CC   DNA.  ATPAC is a new and distinct member of the mdr family of
CC   ATP binding cassette (ABC) transporters.  ATPAC encodes a functional
CC   and robust p-glycoprotein which is inducible upon exposure of the
CC   plant to NPPB.  Expression cassettes containing the ATPAC nucleotide
CC   sequence can be used to produce plants with enhanced resistance to
CC   xenobiotic compounds by in vitro transformation.  Transgenic plants
CC   containing the ATPAC gene are useful for improving stress resistance
CC   or tolerance in plants.  Both over- and under-expression of the ATPAC
CC   gene in transgenic plants have uses in the research of herbicides and
CC   other xenobiotic compounds.
XX
SQ   Sequence 14000 BP; 4465 A; 2277 C; 2321 G; 4937 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 21;  Length 14000;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   385 aatgaaaaaaaaaaaaaat 403
         |||||||||||||||||||
```

```
Db   5825 AATGAAAAAAAAAAAAAAT 5807

RESULT 108
V70401
ID   V70401 standard; DNA; 15462 BP.
XX
AC   V70401;
XX
DT   01-MAR-1999  (first entry)
XX
DE   Human parainfluenza virus type 3 genome..
XX
KW   HPIV-3; vaccine; ds.
XX
OS   Human parainfluenza virus type 3.
XX
FH   Key             Location/Qualifiers
FT   prim_transcript 1702
FT                   /*tag= a
FT                   /note= "end of NP gene"
FT   prim_transcript 1705..3724
FT                   /*tag= b
FT                   /codon_start= 1784..1786
FT                   /note= "P gene"
FT   prim_transcript 3728..4875
FT                   /*tag= c
FT                   /note= "M gene"
FT   prim_transcript 4879..6729
FT                   /*tag= d
FT                   /codon_start= 5072..5074
FT                   /note= "F gene"
FT   prim_transcript 6733..8620
FT                   /*tag= e
FT                   /codon_start= 6806..6808
FT                   /note= "HN start"
FT   prim_transcript 8624..15418
FT                   /*tag= f
FT                   /codon_start= 8646..8648
FT                   /note= "L gene"
FT   misc_feature    15419..15462
FT                   /*tag= g
FT                   /note= "trailer"
XX
PN   WO9850405-A1.
XX
PD   12-NOV-1998.
XX
PF   06-MAY-1998;   98WO-US09270.
XX
PR   07-MAY-1997;   97US-0045805.
```



```
XX
PA   (CLEV-) CLEVELAND CLINIC FOUND.
XX
PI   Banerjee AK,  Hoffman MA;
XX
DR   WPI; 1998-610318/51.
XX
PT   New human parainfluenza virus (HPIV) clone, and - methods of clone
PT   production and introduction of site-specific mutations
XX
PS   Disclosure; Fig 1a-cc; 54pp; English.
XX
CC   This is the full-length, genomic sequence of human parainfluenza
CC   virus type 3 (HPIV-3). In addition to leader and trailer sequences,
CC   the HPIV-3 clone contains sequences encoding the N, P, M, F, HN and
CC   L proteins, as well as cis-acting element.  A claimed recombinant
CC   HPIV clone comprises a nucleotide sequence encoding a positive
CC   sense, antigenome of HPIV, and an RNA polymerase promoter
CC   operatively linked to the sequence.  Also claimed is: a host cell
CC   for producing HPIV recombinants, comprising a HPIV clone as above;
CC   and support clones consisting of a nucleotide sequence encoding
CC   HPIV L and P proteins, where such sequences may be on the same or
CC   separate support clones.  Generation of recombinant HPIV is
CC   especially useful in producing infectious HPIV-3.  The clones also
CC   form a method of producing the clones themselves, through
CC   transfecting the host cells with the appropriate clones (claimed),
CC   and a method of introducing site-specific mutations into the HPIV
CC   genome (claimed).  Production of recombinant HPIV clones enables
CC   analysis of HPIV transcription and replication in a cell-free
CC   system.  Also, preparation of recombinant HPIV-3 is made possible
CC   where it had not previously been, in addition to systems permitting
CC   the introduction of site-specific mutations into the genome of
CC   HPIV-3, which can be identified in order to produce attenuated
CC   HPIV useful for developing a live vaccine.
XX
SQ   Sequence 15462 BP; 5931 A; 2580 C; 2831 G; 4120 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 19;  Length 15462;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4788 agggcaaaagaaaatctca 4806
           ||||||||||||||||||| 
Db   2593 agggcaaaagaaaatctca 2611


RESULT 109
V59470
ID   V59470 standard; DNA; 15512 BP.
XX
```



```
AC   V59470;
XX
DT   13-JAN-1999  (first entry)
XX
DE   Human ryanodin receptor hRyR3 DNA.
XX
KW   Ryanodin receptor type 3; hRyR3; detection; antibody; disease; brain;
KW   smooth muscle; calcium release; intracellular calcium regulation; tumour;
KW   locomotor activity; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             87..14687
FT                   /*tag= a
FT                   /product= "Ryanodin receptor"
XX
PN   DE19722317-C1.
XX
PD   08-OCT-1998.
XX
PF   28-MAY-1997;   97DE-1022317.
XX
PR   28-MAY-1997;   97DE-1022317.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM INT GMBH.
XX
PI   Barsoumian EL,  Hakamata Y,  Nishimura S;
XX
DR   WPI; 1998-522315/45.
DR   P-PSDB; W77410.
XX
PT   New human ryanodin receptor type 3 polypeptide - and corresponding
PT   chimeric polypeptide and DNA, useful in diagnosis and therapy of
PT   associated diseases
XX
PS   Claim 1; Fig 8a-h; 28pp; German.
XX
CC   This sequence encodes a human ryanodin receptor type 3 (hRyR3). This
CC   nucleic acid could be used as a hybridisation probe to detect the
CC   nucleic acid, the polypeptide can be used to produce mono- or
CC   polyclonal antibodies to detect the polypeptide. The polypeptide or
CC   nucleic acid could be used to detect activators and/or inhibitors of
CC   hRyR3, especially in an assay based on determination of ion flux or
CC   intracellular ion concentration. Antibodies or nucleic acid can also be
CC   used for diagnosis of pathological states in tissues by detecting the
CC   presence, over-expression or under-expression of the polypeptide and the
CC   polypeptide or nucleic acid to treat hRyR3-associated diseases. RyR3 is
CC   involved in the release of calcium in the brain and smooth muscle and in
CC   intracellular calcium regulation and may be detectable in pathologically
CC   modified human tissue, e.g. brain tumours. RyR3 deficient mutant mice
```



```
CC   show increased locomotor activity.
XX
SQ   Sequence 15512 BP; 4223 A; 3655 C; 4010 G; 3624 T; 0 other;


  Query Match             0.3%;  Score 19;  DB 19;  Length 15512;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1079 tcatgggcgtgtttgatga 1097
           ||||||||||||||||||| 
Db   5290 tcatgggcgtgtttgatga 5308


RESULT 110
V57274
ID   V57274 standard; DNA; 25464 BP.
XX
AC   V57274;
XX
DT   12-NOV-1998  (first entry)
XX
DE   Human flavin-containing mono-oxygenase isoform x genomic sequence.
XX
KW   Human; flavin-containing mono-oxygenase; FMO; BAC; STS; probe; FISH;
KW   bacterial artificial chromosome; sequence tagged site; chromosome;
KW   fluorescence in situ hybridisation; homology; primer; allele; mutation;
KW   detection; diagnosis; juvenile wide-angle primary glaucoma; stress; drug;
KW   apoptosis; central nervous system; degeneration; oxidation; metabolism;
KW   xenobiotic; carcinogen; biosynthesis; biodegradation; detoxification;
KW   pesticide; herbicide; alkaloid; food additive; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   prim_transcript 2001..25180
FT                   /*tag= a
FT   CDS             2006..25180
FT                   /*tag= b
FT                   /product= "hFMOx"
FT                   /note= "sequence contains introns"
FT   exon            2001..2138
FT                   /*tag= c
FT                   /number= 1
FT   intron          2139..6960
FT                   /*tag= d
FT                   /number= 1
FT   exon            6961..7149
FT                   /*tag= e
FT                   /number= 2
FT   intron          7150..10143
```

```
FT                   /*tag= f
FT                   /number= 2
FT   exon            10144..10306
FT                   /*tag= g
FT                   /number= 3
FT   intron          10307..11412
FT                   /*tag= h
FT                   /number= 3
FT   exon            11413..11555
FT                   /*tag= i
FT                   /number= 4
FT   intron          11556..13346
FT                   /*tag= j
FT                   /number= 4
FT   exon            13347..13546
FT                   /*tag= k
FT                   /number= 5
FT   intron          13547..15696
FT                   /*tag= l
FT                   /number= 5
FT   exon            15697..16052
FT                   /*tag= m
FT                   /number= 6
FT   intron          16053..17929
FT                   /*tag= n
FT                   /number= 6
FT   exon            17930..18002
FT                   /*tag= o
FT                   /number= 7
FT   intron          18003..24837
FT                   /*tag= p
FT                   /number= 7
FT   exon            24838..25180
FT                   /*tag= q
FT                   /number= 8
XX
PN   WO9824914-A1.
XX
PD   11-JUN-1998.
XX
PF   05-DEC-1997;   97WO-FR02226.
XX
PR   06-DEC-1996;   96FR-0015032.
XX
PA   (GEST ) GENSET.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bach JF,  Blumenfeld M,  Garchon HJ,  Tchoumakov I;
XX
DR   WPI; 1998-333333/29.
DR   P-PSDB; W49700.
```



```
XX
PT   Nucleic acid encoding iso-forms 2 and x of human flavin-containing
PT   mono-oxygenase - useful for diagnosis, treatment and prevention of,
PT   e.g. wide-angle primary glaucoma and for synthesis and degradation
PT   of compounds
XX
PS   Claim 2; Page 69-85; 163pp; French.
XX
CC   This sequence represents the genomic clone encoding the human
CC   flavin-containing mono-oxygenase isoform x (hFMOx).  The sequence was
CC   isolated from a bacterial artificial chromosome (BAC) library covering
CC   the complete human genome.  Clone BAC 123H04M was isolated using STS
CC   probes and markers based on human chromosome 1 (the gene's location
CC   was confirmed as being in the region 1q23-q25 by fluorescence in situ
CC   hybridisation).  The BAC clone was digested with restriction enzymes.
CC   The fragments were subcloned and the inserts sequenced.  The human FMO
CC   sequence was identified by homology to known FMO sequences. Fragments
CC   of the FMOx sequence can be used as primers or probes specific for
CC   particular alleles, especially the G.1263mac.A mutation.  Detection of
CC   this mutation allows diagnosis of a predisposition to FMO-associated
CC   diseases, particularly juvenile wide-angle primary glaucoma (WAPG),
CC   but more generally any condition involving altered response to
CC   oxidative stress, including toxic and degenerative disorders,
CC   particularly where apoptosis is involved, or central nervous system
CC   degeneration.  Other FMO-related disorders are excessive or inadequate
CC   oxidation/metabolism of drugs, toxic xenobiotics or carcinogens.
CC   The proteins and cells expressing them can also be used for the
CC   biosynthesis/biodegradation, particularly detoxification, of xenobiotics
CC   (pesticides, herbicides, alkaloids and pharmaceuticals) and synthesis
CC   of pharmaceuticals, food additives, pesticides and herbicides.
XX
SQ   Sequence 25464 BP; 8183 A; 4947 C; 4984 G; 7350 T; 0 other;
```

```
  Query Match             0.3%;  Score 19;  DB 19;  Length 25464;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgctt 407
           ||||||||||||||||||
Db    4754 aaaaaaaaaaaaaatgctt 4772


RESULT 111
X12956/c
ID   X12956 standard; DNA; 25580 BP.
XX
AC   X12956;
XX
DT   19-MAR-1999  (first entry)
XX
```



```
DE   Enterococcus faecalis genome contig SEQ ID NO:19.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 341-354; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 25580 BP; 9236 A; 3779 C; 5253 G; 7286 T; 26 other;
```

```
  Query Match             0.3%;  Score 19;  DB 20;  Length 25580;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy    2625 tgataaatcttttgatgaa 2643
           ||||||||||||||||||
Db   19880 TGATAAATCTTTTGATGAA 19862


RESULT 112
V21210/c
ID   V21210 standard; DNA; 58407 BP.
XX
AC   V21210;
XX
DT   10-NOV-1998  (first entry)
XX
DE   Methanococcus jannaschii large circular extrachromosomal element.
XX
KW   Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW   genome; autotrophic; extrachromosomal element; identification; ds.
XX
OS   Methanococcus jannaschii.
XX
PN   WO9807830-A2.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-US14900.
XX
PR   22-AUG-1996;   96US-0024428.
XX
PA   (GENO-) INST GENOMIC RES.
PA   (UNII ) UNIV ILLINOIS POUND.
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Bult CJ,  Smith HO,  Venter JC,  White OR,  Woese CR;
XX
DR   WPI; 1998-169145/15.
XX
PT   Complete genome sequence of methano-genic archaeon, Methanoccoccus
PT   jannaschii  - useful in identification of M. jannaschii genome
PT   fragment
XX
PS   Claim 13; Page 585-600; 614pp; English.
XX
CC   The present sequence represents the large circular extrachromosomal
CC   element sequence of the Methanococcus jannaschii circular chromosome. The
CC   present invention describes M. jannaschii open reading frames from the
CC   genome sequence. The invention also describes a computer based system
CC   for identifying fragments of the M. jannaschii genome that are
CC   homologous to target nucleotide sequences, comprising: (a) data storage
CC   means comprising the nucleotide sequence of the 1664976, 58407 or 16550
CC   bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at
```

```
CC   least 99.9% identical to it; (b) search means for comparing a target
CC   sequence to the nucleotide sequence of the data storage means to
CC   identify a homologous sequence, and (c) retrieval means for obtaining
CC   the homologous sequence. The method, which is based on whole genome
CC   random sequencing of an autotrophic archaeon M. jannaschii, the genome
CC   of which consists of 3 physically distinct elements, a large circular
CC   chromosome (the 1664976 bp sequence given in V21209), a large circular
CC   extra-chromosomal element (the 58407 bp sequence given in V21210), and a
CC   small circular extra-chromosomal element (the 16550 bp sequence given in
CC   V21211), can be used in the identification of M. jannaschii genome
CC   fragment.
XX
SQ   Sequence 58407 BP; 20163 A; 9034 C; 7447 G; 21761 T; 2 other;


  Query Match             0.3%;  Score 19;  DB 19;  Length 58407;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 attttaagcaactccatag 23
           ||||||||||||||||||
Db   20916 ATTTTAAGCAACTCCATAG 20898


RESULT 113
X57351
ID   X57351 standard; DNA; 68940 BP.
XX
AC   X57351;
XX
DT   16-AUG-1999  (first entry)
XX
DE   Human chromosome 6q27 tumour suppressor gene.
XX
KW   Tumour suppressor gene; chromosome 6q27; cancer; therapy; human; D6S297;
KW   breast cancer; ovarian cancer; lymphoid cancer; marker; D6S193; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            414..549
FT                   /*tag= a
FT   polyA_site      complement (719..724)
FT                   /*tag= b
FT   repeat_region   1126..1220
FT                   /*tag= c
FT                   /note= "MIR repeat family"
FT   variation       replace(1128,T)
FT                   /*tag= d
FT                   /note= "putative polymorphism, C in 52b3, T in
FT                         pDJ84g15"
```



```
FT   variation       replace(1909,C)
FT                   /*tag= e
FT                   /note= "putative polymorphism, A in 52b3, C in
FT                         "pDJ84g15"
FT   variation       replace(1920,T)
FT                   /*tag= f
FT                   /note= "putative polymorphism, C in 52b3, T in
FT                         pDJ84g15"
FT   exon            complement (2310..2482)
FT                   /*tag= g
FT   exon            2749..2843
FT                   /*tag= h
FT   variation       replace(2984,T)
FT                   /*tag= i
FT                   /note= "putative polymorphism, C in 52b3, T in
FT                         pDJ84g15"
FT   exon            complement (3083..3201)
FT                   /*tag= j
FT                   /note= "trapped exon"
FT   exon            3654..3793
FT                   /*tag= k
FT   promoter        complement (4483..4532)
FT                   /*tag= l
FT   promoter        complement (4615..4664)
FT                   /*tag= m
FT   promoter        4854..4903
FT                   /*tag= n
FT   promoter        4878..4927
FT                   /*tag= o
FT   exon            4912..5093
FT                   /*tag= p
FT   misc_feature    complement (5398..5563)
FT                   /*tag= q
FT                   /note= "match: D6S193 microsatellite"
FT   promoter        complement (5558..5607)
FT                   /*tag= r
FT   promoter        5767..5816
FT                   /*tag= s
FT   exon            6072..6134
FT                   /*tag= t
FT   exon            complement (6395..6438)
FT                   /*tag= u
FT                   /note= "trapped exon"
FT   exon            complement (6395..6456)
FT                   /*tag= v
FT   polyA_site      complement (6887..6892)
FT                   /*tag= w
FT   exon            complement (7237..7434)
FT                   /*tag= x
FT   exon            complement (8055..8165)
FT                   /*tag= y
```



```
FT   exon            complement (8414..8447)
FT                   /*tag= z
FT   exon            8530..8695
FT                   /*tag= aa
FT   exon            complement (8543..8593)
FT                   /*tag= ab
FT                   /note= "trapped exon"
FT   exon            9351..9678
FT                   /*tag= ac
FT                   /note= "trapped exon"
FT   exon            complement (9467..9583)
FT                   /*tag= ad
FT   exon            9627..9678
FT                   /*tag= ae
FT   promoter        complement (9636..9685)
FT                   /*tag= af
FT   exon            complement (10423..10461)
FT                   /*tag= ag
FT   exon            complement (10669..10761)
FT                   /*tag= ah
FT   exon            complement (10706..10761)
FT                   /*tag= ai
FT   repeat_region   11067..11368
FT                   /*tag= aj
FT                   /note= "Alu repeat family"
FT   exon            11264..11315
FT                   /*tag= ak
FT   polyA_site      complement (12270..12275)
FT                   /*tag= al
FT   promoter        complement (12641..12690)
FT                   /*tag= am
FT   exon            13800..14074
FT                   /*tag= an
FT                   /note= "trapped exon"
FT   exon            complement (14197..14400)
FT                   /*tag= ao
FT   repeat_region   complement (14496..14601)
FT                   /*tag= ap
FT                   /note= "L1 repeat family"
FT   exon            complement (14607..14860)
FT                   /*tag= aq
FT                   /note= "trapped exon"
FT   promoter        14930..14979
FT                   /*tag= ar
FT   variation       replace(15091,G)
FT                   /*tag= as
FT                   /note= "putative polymorphism, A in 52b3 and
FT                         pDJ84g15, G in 255h6"
FT   polyA_site      complement (15609..15614)
FT                   /*tag= at
FT   repeat_region   16174..16451
```



```
FT                   /*tag= au
FT                   /note= "Alu repeat family"
FT   exon            16509..16559
FT                   /*tag= av
FT   exon            complement (16867..16939)
FT                   /*tag= aw
FT                   /note= "trapped exon"
FT   exon            16867..17013
FT                   /*tag= ax
FT                   /note= "trapped exon"
FT   exon            17288..17352
FT                   /*tag= ay
FT   exon            complement (17427..17606)
FT                   /*tag= az
FT   promoter        17647..17696
FT                   /*tag= ba
FT   promoter        complement (18367..18416)
FT                   /*tag= bb
FT   exon            complement (19301..19334)
FT                   /*tag= bc
FT   repeat_region   complement (19541..19683)
FT                   /*tag= bd
FT                   /note= "MER20 repeat family"
FT   repeat_region   19688..19975
FT                   /*tag= be
FT                   /note= "Alu repeat family"
FT   exon            20108..20226
FT                   /*tag= bf
FT   variation       replace(20729..20767)
FT                   /*tag= bg
FT                   /note= "putative polymorphism (TAT13) in 52b3"
FT   repeat_region   complement (20770..20878)
FT                   /*tag= bh
FT                   /note= "Alu repeat family"
FT   promoter        21292..21341
FT                   /*tag= bi
FT   promoter        21450..21499
FT                   /*tag= bj
FT   promoter        complement (22392..22441)
FT                   /*tag= bk
FT   exon            22856..22956
FT                   /*tag= bl
FT   exon            complement (23196..23233)
FT                   /*tag= bm
FT   polyA_site      complement (23288..23293)
FT                   /*tag= bn
FT   variation       replace(24182..24186)
FT                   /*tag= bo
FT                   /note= "putative polymorphism (T)5 in pDJ84g15"
FT   promoter        complement (24244..24293)
FT                   /*tag= bp
```

```
FT   promoter        24266..24315
FT                   /*tag= bq
FT   promoter        complement(24600..24649)
FT                   /*tag= br
FT   exon            30550..31665
FT                   /*tag= bs
FT   exon            30550..30863
FT                   /*tag= bt
FT   polyA_site      32483..32488
FT                   /*tag= bu
FT   repeat_region   32483..32488
FT                   /*tag= bv
FT   repeat_region   33639..33870
FT                   /*tag= bw
FT                   /note= "Alu repeat family"
FT   exon            complement (33782..33867)
FT                   /*tag= bx
FT   repeat_region   complement (34638..34996)
FT                   /*tag= by
FT                   /note= "TIGGER2 repeat family"
FT   exon            complement (34879..34941)
FT                   /*tag= bz
```

```
  Query Match             0.3%;  Score 19;  DB 20;  Length 68940;
  Best Local Similarity   100.0%;  Pred. No. 78;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgctt 407
          ||||||||||||||||||
Db  64556 aaaaaaaaaaaaaatgctt 64574
```

```
RESULT 114
Z86967/c
ID   Z86967 standard; DNA; 162450 BP.
XX
AC   Z86967;
XX
DT   16-MAY-2000  (first entry)
XX
DE   Retinoblastoma binding protein-7 genomic DNA sequence.
XX
KW   RBP-7; retinoblastoma binding protein-7; abnormal cell proliferation;
KW   diagnosis; therapy; cell differentiation; thyroid hyperplasia; psoriasis;
KW   benign prostate hypertrophy; cancer; sarcoma; neoplasm; leukaemia;
KW   lymphoma; ds.
XX
OS   Homo sapiens.
XX
PN   WO200000607-A1.
XX
```



```
PD   06-JAN-2000.
XX
PF   30-JUN-1999;   99WO-IB01242.
XX
PR   30-JUN-1998;   98US-0091315.
PR   10-DEC-1998;   98US-0111909.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L;
XX
DR   WPI; 2000-117170/10.
XX
PT   Novel nucleic acid and polymorphic markers used for diagnosis of
PT   diseases, especially those involving abnormal cell proliferation and
PT   differentiation -
XX
PS   Claim 1; Page 118-163; 223pp; English.
XX
CC   This sequence represents the retinoblastoma binding protein-7 (RBP-7)
CC   genomic sequence of the invention. The RBP-7 coding sequence and
CC   regulatory sequences are useful for the recombinant production of the
CC   protein and for expressing heterologous nucleic acids. Primers and probes
CC   derived from the RBP-7 nucleotide sequence (e.g. Z87035-Z87099) are
CC   useful for DNA amplification and detection methods. RBP-7 biallelic
CC   markers (see Z86993-Z87034) are useful for diagnosis of diseaae related
CC   to alteration in the regulation or in the coding regions of the RBP-7
CC   gene and for prognosis/diagnosis of an eventual treatment with
CC   therapeutic agents, especially agents acting on pathologies involving
CC   abnormal cell proliferation and/or differentiation, these include
CC   thyroid hyperplasia, psoriasis, benign prostate hypertrophy, cancers,
CC   including breast cancer, sarcomas and other neoplasms, bladder cancer,
CC   colon cancer, lung cancer, prostate cancer, various leukaemias, and
CC   lymphomas. RBP-7 antibodies are useful as diagnostic agents.
XX
SQ   Sequence 162450 BP; 45465 A; 30661 C; 32637 G; 53673 T; 14 other;
```

```
  Query Match             0.3%;  Score 19;  DB 21;  Length 162450;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2642 aattcaaaaatggcaaaaa 2660
          ||||||||||||||||||
Db 122668 AATTCAAAAATGGCAAAAA 122650
```

```
RESULT 115
V73529/c
ID   V73529 standard; DNA; 18 BP.
XX
```



```
AC   V73529;
XX
DT   22-MAR-2000  (first entry)
XX
DE   H. pylori cag A primer cagRN1.
XX
KW   PCR primer; probe; vacA; cagA; detection; vacuolating toxin; VDG;
KW   virulence determinant gene; cytotoxin-associated gene; allele-specific;
KW   infectivity; pathogenicity; gastritis; gastric; duodenal; ulcer;
KW   adenocarcinoma; mucosa-associated lymphoid tissue lymphoma; therapy;
KW   S region; S1a; S1b; S1c; S2; M region; M1; M2; ss.
XX
OS   Helicobacter pylori.
XX
PN   WO9816650-A2.
XX
PD   23-APR-1998.
XX
PF   10-OCT-1997;   97WO-EP05614.
XX
PR   16-OCT-1996;   96EP-0870131.
PR   09-SEP-1997;   97EP-0870133.
XX
PA   (INNO-) INNOGENETICS NV.
PA   (DDLD-) DDL BV.
XX
PI   Quint W,  Van Doorn L;
XX
DR   WPI; 1998-251300/22.
XX
PT   Method for detecting and/or typing Helicobacter pylori strains -
PT   comprises use of primers and probes based on vacA and cagA gene
XX
PS   Claim 2; Page 44; 122pp; English.
XX
CC   This invention describes a novel method for the detection and/or typing
CC   of Helicobacter pylori strains present in a sample using PCR primers and
CC   probes to detect regions of the vacuolating toxin (vacA) gene and other
CC   virulence determinant genes (VDG) e.g. the cytotoxin-associated (cagA)
CC   gene. The method allows the typing and allele-specific detection of a
CC   strain according to the VDG alleles present in that particular H. pylori
CC   strain. The virulence determinant genes are the genetic elements involved
CC   in enabling, determining, and marking the infectivity and/or
CC   pathogenicity of the H. pylori strain. The method provides a way of
CC   detecting H. pylori strains in a sample with respect to the development
CC   of chronic active gastritis, gastric and duodenal ulcers, gastric
CC   adenocarcinomas, mucosa-associated lymphoid tissue lymphomas, and/or
CC   determining eradication therapy. V73508-V73546 represent PCR primers
CC   and probes used in the detection of the H. pylori vacA and cagA genes.
CC   The primers and probes are used especially to detect the vacA S regions
CC   S1a/b/c and S2 and the M regions M1 and M2 which are represented in
```



```
CC   V73547-V73785.
XX
SQ   Sequence 18 BP; 4 A; 2 C; 3 G; 9 T; 0 other;
```

```
  Query Match             0.3%;  Score 18;  DB 19;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 3.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    764 ccaaaaagaatcagtatt 781
          ||||||||||||||||||
Db     18 CCAAAAAGAATCAGTATT 1
```

```
RESULT 116
V65775
ID   V65775 standard; DNA; 18 BP.
XX
AC   V65775;
XX
DT   18-JAN-1999  (first entry)
XX
DE   Helicobacter pylori cagA gene primer DZ3.
XX
KW   Porin; porA gene; cytotoxin; infection; diagnosis; vaccine;
KW   Campylobacter jejuni; cagA gene; PCR; primer; ss.
XX
OS   Synthetic.
OS   Helicobacter pylori.
XX
PN   WO9842842-A1.
XX
PD   01-OCT-1998.
XX
PF   25-MAR-1998;   98WO-CA00272.
XX
PR   25-MAR-1997;   97US-0041200.
XX
PA   (CNDG ) CANADA MIN HEALTH & WELFARE.
XX
PI   Bacon DJ,  Bolla J,  Johnson WM,  Rodgers P;
XX
DR   WPI; 1998-532008/45.
XX
PT   New nucleic acid encoding the porA porin from Campylobacter jejuni -
PT   useful for detecting C. jejuni infections in humans and animals
XX
PS   Example 2; Page 51; 131pp; English.
XX
CC   Primers DZ3 and R009 (see V65775-76) are specific for the cagA
CC   gene of Helicobacter pylori.  These primers, along with other
```

```
CC   primers specific for the Escherichia coli VT1 gene (see V65771-72),
CC   the E. coli VT2 gene (see V65773-74), or the H. pylori vacA gene
CC   (see V65777-78), were used in the PCR amplification of
CC   Campylobacter jejuni strain 2483 genomic DNA.  No amplicons were
CC   generated.  The C. jejuni strain 2483 porin (see W79959) shows
CC   homology to another isolated C. jejuni cytotoxic porin protein,
CC   to the C. jejuni major outer membrane protein, and to the 45 and 51
CC   kDa outer membrane proteins of Wolinella recta.  The isolated
CC   porA gene (see V65769) of strain 2483, and porin polypeptides, can
CC   be used in methods for detecting C. jejuni infections in humans
CC   and animals.  Immunogenically active peptides of porin are used in
CC   vaccines.
XX
SQ   Sequence 18 BP; 10 A; 1 C; 5 G; 2 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 19;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 3.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     521 agtaaggagaaacaatga 538
           ||||||||||||||||||
Db       1 agtaaggagaaacaatga 18


RESULT 117
Q75761/c
ID   Q75761 standard; DNA; 21 BP.
XX
AC   Q75761;
XX
DT   04-AUG-1995  (first entry)
XX
DE   Reverse transcription primer used in cDNA analysis technique.
XX
KW   Analysis; gene expression; reverse transcription; primer; cDNA;
KW   aggregate; restriction enzyme; ss.
XX
OS   Synthetic.
XX
PN   JP06303997-A.
XX
PD   01-NOV-1994.
XX
PF   16-APR-1993;   93JP-0112515.
XX
PR   16-APR-1993;   93JP-0112515.
XX
PA   (NITE ) NIPPON TELEGRAPH & TELEPHONE CORP.
XX
DR   WPI; 1995-018287/03.
```



```
XX
PT   Analysis of cDNA and gene expression - by amplification of mRNA
PT   followed by digestion with restriction enzymes
XX
PS   Disclosure; Page 8; 11pp; Japanese.
XX
CC   A method for the analysis of cDNA comprises (a) preparing an
CC   aggregate of double-stranded cDNAs by using an aggregate of mRNAs
CC   and a plural type of labelled reverse transcription primers
CC   (GENESEQ files Q75547-Q75798) and using the aggregate of mRNAs as the
CC   template for each reverse transcription primer; (b) digesting each of
CC   the prepared aggregates of the double-stranded cDNAs with restriction
CC   enzyme and; (c) electrophoresing the digested aggregate of cDNAs in
CC   seperate lanes. The method can be used to analyse gene expression
CC   rapidly and easily.
XX
SQ   Sequence 21 BP; 1 A; 1 C; 0 G; 19 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 16;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      21 AATGAAAAAAAAAAAAAA 4


RESULT 118
T61445/c
ID   T61445 standard; DNA; 40 BP.
XX
AC   T61445;
XX
DT   30-OCT-1997  (first entry)
XX
DE   Nucleotide anchored heeled oligo-dT primer RP6.0.
XX
KW   Reverse transcription; restriction enzyme; adapter; PCR;
KW   amplification; isolation; T-cell; activation; expression; ss.
XX
OS   Synthetic.
XX
PN   WO9705286-A1.
XX
PD   13-FEB-1997.
XX
PF   30-JUL-1996;   96WO-US12468.
XX
PR   01-AUG-1995;   95US-0510032.
XX
```



```
PA   (UYYA ) UNIV YALE.
XX
PI   Prashar Y,  Weissman S;
XX
DR   WPI; 1997-145720/13.
XX
PT   Isolating DNA complementary to mRNA in a sample - by reverse
PT   transcription, restriction enzyme cutting, ligation to a Y-shaped
PT   adapter and selective PCR amplification
XX
PS   Claim 3; Page 32; 46pp; English.
XX
CC   The sequences given in T61442 to T61453 are claimed and are used
CC   to exemplify the new method for isolating DNA in a nucleic acid
CC   sample.  RNA expression during early T-cell activation was studied
CC   as test system.
XX
SQ   Sequence 40 BP; 7 A; 8 C; 4 G; 21 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 18;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaatgct 406
           ||||||||||||||||||
Db      34 AAAAAAAAAAAAATGCT 17


RESULT 119
Z25103/c
ID   Z25103 standard; DNA; 40 BP.
XX
AC   Z25103;
XX
DT   13-DEC-1999  (first entry)
XX
DE   Human ischaemic heart disease associated protein primer #3.
XX
KW   Human; ischaemic heart disease; ischaemia; heart tissue; detection;
KW   ischaemic cardiomyopathy; diagnosis; primer; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9949062-A1.
XX
PD   30-SEP-1999.
XX
PF   26-MAR-1999;   99WO-US06662.
XX
```



```
PR   26-MAR-1998;   98US-0079377.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Prashar Y,  Fan W;
XX
DR   WPI; 1999-580453/49.
XX
PT   New isolated nucleic acids associated with ischaemic heart tissue, used
PT   to develop products for treating e.g. ischaemic heart disease and
PT   ischaemic cardiomyopathy
XX
PS   Example 1; Page 23; 51pp; English.
XX
CC   The present sequence represents a primer for human ischaemic heart
CC   disease associated protein (I), which is used in an example from the
CC   present invention. The (I) nucleotide sequence is up-regulated in
CC   ischaemic heart tissue. The (I) proteins can serve as a target for agents
CC   that can be used to modulate the expression or activity of the protein,
CC   e.g. agents may be identified which modulate biological processes
CC   associated with ischaemic injury to the heart such as chronic ischaemic
CC   heart disease and ischaemic cardiomyopathy. Agents may also be identified
CC   which modulate the biological process associated with recovery to
CC   ischaemic injury to the heart. They can also be used for isolated
CC   binding partners that bind to the protein. The proteins also provide a
CC   novel target for screening of synthetic small molecules and
CC   combinatorial or naturally occurring compound libraries to discover
CC   novel therapeutics to regulate heart function. The products can also be
CC   used for detection, diagnosis and production of transgenic animals.
XX
SQ   Sequence 40 BP; 7 A; 8 C; 4 G; 21 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaatgct 406
           ||||||||||||||||||
Db      34 AAAAAAAAAAAAATGCT 17


RESULT 120
X36067/c
ID   X36067 standard; DNA; 40 BP.
XX
AC   X36067;
XX
DT   15-JUL-1999  (first entry)
XX
DE   Primer used to synthesise cDNA associated with neutrophil activation.
```

XX
KW Neutrophil; differentially expressed; bacteria; activation;
KW nucleic acid probe; amplification primer; hybridisation assay;
KW pathogen; primer; ss.
XX
OS Synthetic.
XX
PN WO9920642-A2.
XX
PD 29-APR-1999.
XX
PF 21-OCT-1998; 98WO-US22171.
XX
PR 21-OCT-1997; 97US-0062564.
XX
PA (UYYA) UNIV YALE.
XX
PI Goguen J, Newburger P, Prashar Y, Weissman S, Yamaga S;
PI Yerramilli S;
XX
DR WPI; 1999-288271/24.
XX
PT Unique nucleic acid sequence
XX
PS Example 1; Page 25; 74pp; English.
XX
CC The specification describes neutrophil nucleic acid sequences
CC (X35979-X36064) which correspond to mRNA, and therefore to genes. These
CC sequences are differentially expressed upon exposure of the neutrophils
CC to virulent or avirulent bacteria. The nucleic acid sequences can be
CC used as nucleic acid probes to screen cDNA and genomic libraries to
CC select similar sequences; to generate amplification primers; and to
CC identify adjacent and flanking sequences in the genome. The sequences
CC can be used, e.g. in hybridisation assays, to detect activated
CC neutrophils or neutrophils previously exposed to virulent or avirulent
CC bacteria. The sequences can also be used as probes to monitor the
CC expression of at lest one differentially expressed neutrophil gene in
CC a method to identify a therapeutic or prophylactic agent that modulates
CC the response of a neutrophil to a pathogen. The present primer is used
CC in the course of the invention.
XX
SQ Sequence 40 BP; 7 A; 8 C; 4 G; 21 T; 0 other;

Query Match 0.3%; Score 18; DB 20; Length 40;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 389 aaaaaaaaaaaaaatgct 406
||||||||||||||||||
Db 34 AAAAAAAAAAAAAATGCT 17

RESULT 121
Z51197/c
ID Z51197 standard; DNA; 40 BP.
XX
AC Z51197;
XX
DT 06-JUN-2000 (first entry)
XX
DE Primer RP6.0, for cDNA synthesis of ischaemic heart mRNA.
XX
KW ischaemia; heart disease; human; PGP980; cardiomyopathy; cardiant;
KW modulator; treat; prevention; diagnosis; primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200011942-A1.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20015.
XX
PR 01-SEP-1998; 98US-0098683.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Einstein R;
XX
DR WPI; 2000-237720/20.
XX
PT Isolated nucleic acids encoding a proteins that is differentially
PT expressed during ischemic heart disease and ischemic cardiomyopathy,
PT useful for preventing, diagnosing and treating cardiac disease
XX
PS Example 1; Page 56; 61pp; English.
XX
CC The present sequence is the primer RP6.0, used for cDNA synthesis of the
CC differentially expressed protein PGP980, from ischaemic heart mRNA. It
CC consists of two-base anchored oligo (dT) primer and is used with other
CC components of the first-strand cDNA synthesis. The protein PGP980, is
CC expressed in heart, brain, lung, placenta, skeletal muscle, kidney and
CC pancreas and has cardiant activity. The DNA may be used as probes to
CC detect and quantify homologous sequences in samples. PGP980 sequences can
CC be used to identify modulators, that can be used for treatment, diagnosis
CC and prevention of diseases associated with inappropriate expression and
CC activity of PGP980, such as ischaemic heart disease and ischaemic
CC cardiomyopathy.
XX
SQ Sequence 40 BP; 7 A; 8 C; 4 G; 21 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 389 aaaaaaaaaaaaaatgct 406
||||||||||||||||||
Db 34 AAAAAAAAAAAAAATGCT 17

RESULT 122
Z59453/c
ID Z59453 standard; DNA; 40 BP.
XX
AC Z59453;
XX
DT 11-APR-2000 (first entry)
XX
DE PCR primer RP6.0 used in a method for gene analysis.
XX
KW PCR primer; selective amplification method; gene expression; analysis;
KW mRNA level comparison; ss.
XX
OS Synthetic.
XX
PN US6010850-A.
XX
PD 04-JAN-2000.
XX
PF 30-JUL-1996; 96US-0688514.
XX
PR 01-AUG-1995; 95US-0510032.
XX
PA (UYYA) UNIV YALE.
XX
PI Prashar Y, Weissman SM;
XX
DR WPI; 2000-105493/09.
XX
PT DNA amplification method useful for studying changes in gene expression
PT between cell populations
XX
PS Example 1; Column 6; 23pp; English.
XX
CC This sequence represents a PCR primer used for the amplification of
CC Jurkat T cell cDNA, in the method of the invention. The invention relates
CC to a method for selectively amplifying DNA, where the DNA has a sequence
CC complementary to a 3' end of an mRNA comprising ligating digested cDNA to
CC a adaptor. The method mediates selective PCR amplification of cDNAs under
CC high stringency. The method can be used for comparing the levels of mRNA

CC expression in two cell populations by selectively amplifying a DNA
CC fragment in a nucleic acid sample from each cell and comparing the
CC amounts of amplified fragments obtained. The comparison can be carried
CC out where at least one of the primers is labelled, preferably with a
CC radiolabel or fluorescent label, and where one of the cell populations is
CC a treated population (the other being a control population).
XX
SQ Sequence 40 BP; 7 A; 8 C; 4 G; 21 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 389 aaaaaaaaaaaaaatgct 406
||||||||||||||||||
Db 34 AAAAAAAAAAAAAATGCT 17

RESULT 123
Z38052/c
ID Z38052 standard; DNA; 40 BP.
XX
AC Z38052;
XX
DT 07-FEB-2000 (first entry)
XX
DE Jurkat cell cDNA amplifying primer RP6.0.
XX
KW T lymphocyte; sterile inflammatory disease; autoimmune disorder; GVHD;
KW immunodeficiency disease; cancer; graft-versus-host-disease; psoriasis;
KW rheumatoid arthritis; asthma; genetic disease; cytokine gene; Jurkat;
KW primer; ss.
XX
OS Synthetic.
XX
PN WO9957130-A1.
XX
PD 11-NOV-1999.
XX
PF 05-MAY-1999; 99WO-US09761.
XX
PR 05-MAY-1998; 98US-0084329.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Prashar Y, Weissman S;
XX
DR WPI; 2000-023547/02.
XX
PT Studying changes in gene expression in T lymphocytes used to identify

```
PT   therapeutic or prophylactic agents which modulate T lymphocyte
PT   populations .
XX
PS   Example 1; Page 33; 78pp; English.
XX
CC   The invention provides methods to identify a therapeutic or prophylactic
CC   agent that modulates a T lymphocyte population found in a subject having
CC   a sterile inflammatory disease, autoimmune disorder, immunodeficiency
CC   disease, cancer or graft-versus-host-disease (GVHD). The methods can be
CC   used to identify therapeutic or prophylactic agents that modulate T
CC   lymphocyte populations in subjects having a sterile inflammatory disease
CC   (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma,
CC   allergic rhinitis, cardiac and renal reperfusion injury, thrombosis,
CC   adult respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease, lymphocytopenias,
CC   autoimmune lymphoproliferative syndrome, synovitis, sarcoidosis,
CC   DiGeorge's syndrome, Heselof syndrome, severe combined immunodeficiency
CC   syndromes, Wiskott-Aldrich syndrome, Ataxia-telangiectasia, Cartilage-
CC   hair hypoplasia, immunodeficiency with thymoma and leukocyte adhesion
CC   deficiency 1), cancer (15 T lymphocyte leukemia/lymphoma's are listed in
CC   the specification), immunodeficiency disease, autoimmune disorder (20
CC   listed in the specification) or a genetic disease. The methods can also
CC   be used to identify an agent that induces the differentiation of
CC   quiescent pre-, cortical or medullary T lymphocytes into a differentiated
CC   T lymphocyte subpopulation. The methods are especially useful for
CC   identifying cytokine genes, genes encoding cell surface receptors and
CC   genes encoding intermediary signaling molecules.
XX
SQ   Sequence 40 BP; 7 A; 8 C; 4 G; 21 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 21;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgct 406
          ||||||||||||||||||
Db     34 AAAAAAAAAAAAATGCT 17


RESULT 124
X34852/c
ID   X34852 standard; DNA; 41 BP.
XX
AC   X34852;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Primer for adapter ligated 3' ends of differentially expressed cDNA.
XX
KW   Differentially expressed sequence; expression profile; LIN+ cell;
```



```
KW   LRH cell; LRH48 cell; LRBRH cell; expression; stem cell gene;
KW   differentiation; proliferation; modulator; disease; immune disorder;
KW   malignancy; lympho-hematopoietic system; PCR primer; ss.
XX
OS   Synthetic.
OS   Mus sp.
XX
PN   WO9910535-A1.
XX
PD   04-MAR-1999.
XX
PF   21-AUG-1998;   98WO-US17283.
XX
PR   22-AUG-1997;   97US-0056861.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Baskaran N,  Liu M,  Weissman SM;
XX
DR   WPI; 1999-243570/20.
XX
PT   New identification of agents that modulate gene expression in stem
PT   cells
XX
PS   Example 1; Page 25; 68pp; English.
XX
CC   PCR primers X34850-54 were used to amplify adapter ligated 3' cDNA
CC   ends of differentially expressed cDNA sequences from LIN+, LRH, LRH48
CC   and LRBRH murine cells. The specification describes a method for
CC   identifying an agent that modulates expression of a stem cell gene
CC   associated with differentiation or proliferation by comparison of gene
CC   expression profiles. A composition comprising the sequences of
CC   X26987-27055 and X34838-45 affixed to a solid support can be used in
CC   the method of the invention. The invention is used to identify agents
CC   for the treatment of diseases and immune disorders, including
CC   malignancies related to disruptions within the lympho-hematopoietic
CC   system.
XX
SQ   Sequence 41 BP; 7 A; 8 C; 4 G; 22 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgct 406
          ||||||||||||||||||
Db     34 AAAAAAAAAAAAATGCT 17


RESULT 125
```



```
X23456/c
ID   X23456 standard; DNA; 41 BP.
XX
AC   X23456;
XX
DT   17-JUN-1999  (first entry)
XX
DE   Human neutrophil cDNA synthesis primer 3.
XX
KW   Neutrophil; gene expression profile; granulocyte; pathogen-exposed;
KW   sterile inflammatory disease; detection; therapeutic agent; human;
KW   expression modulator; pathogenic infection; cell activation; primer;
KW   global transcriptional response; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9910536-A1.
XX
PD   04-MAR-1999.
XX
PF   21-AUG-1998;   98WO-US17284.
XX
PR   22-AUG-1997;   97US-0056844.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Goguen J,  Newburger P,  Prashar Y,  Weissman SM,  Yerramilli SV;
XX
DR   WPI; 1999-204678/17.
XX
PT   Detection of pathogen exposure or sterile inflammatory disease in a
PT   subject  - by comparing gene expression profiles from granulocytes
PT   of the patient and control granulocytes
XX
PS   Example 1; Page 23; 84pp; English.
XX
CC   This invention describes a method for the comparison of gene expression
CC   profiles from granulocytes from a test subject and from pathogen-exposed
CC   or sterile inflammatory disease granulocytes or quiescent granulocytes.
CC   The method is used to detect pathogen exposure or sterile inflammatory
CC   disease in a subject and to identify therapeutic agents that modulate
CC   expression of a gene in response to a pathogenic infection or to
CC   sterile inflammatory disease in a subject. The method tests for global
CC   transcriptional response of granulocytes during cell activation.
XX
SQ   Sequence 41 BP; 7 A; 8 C; 4 G; 22 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
```



```
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgct 406
          ||||||||||||||||||
Db     34 AAAAAAAAAAAAATGCT 17


RESULT 126
V48090
ID   V48090 standard; DNA; 43 BP.
XX
AC   V48090;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Oligonucleotide 43-P.
XX
KW   In situ translation; RNA-protein fusion; binding reagent; antibody;
KW   industrial catalyst; ss; rRNA/DNA hybrid.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   modified_base    43
FT                    /*tag= a
FT                    /note= "Puromycin"
XX
PN   WO9831700-A1.
XX
PD   23-JUL-1998.
XX
PF   14-JAN-1998;   98WO-US00807.
XX
PR   06-NOV-1997;   97US-0064491.
PR   21-JAN-1997;   97US-0035963.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Liu R,  Roberts RW,  Szostak JW;
XX
DR   WPI; 1998-414032/35.
XX
PT   Selection of specific protein by screening protein-RNA fusions
PT   generated in vitro or in situ - useful for, e.g. identifying enzymes
PT   and antibodies with altered properties, potentially useful as
PT   catalysts or for therapy or diagnosis
XX
PS   Disclosure; Page 39; 94pp; English.
XX
CC   The Oligonucleotides V48087, V48089-V48091 and V48096-V48098 and
CC   variations were used to generate RNA-protein fusions. These were used in
```

CC the selection of a specific protein or RNA, by in vitro or in situ
CC translation of candidate RNA molecules to produce RNA-protein fusions,
CC then selecting specific RNA protein fusions. The method is used to
CC select proteins (or DNA encoding them) having altered properties, e.g.
CC for identification of new binding reagents, to identify improved human
CC antibodies or new enzymes. These proteins are potentially useful in
CC diagnosis and therapy, or as industrial catalysts. The methods allow
CC many rounds of selection and amplification to be performed, resulting in
CC enrichment of even very rare molecules and allowing isolation of proteins
CC that bind specifically to almost any compound or catalyse almost any
CC reaction.
XX
SQ Sequence 43 BP; 32 A; 3 C; 6 G; 1 U; 1 other;

Query Match 0.3%; Score 18; DB 19; Length 43;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||:||||||||||||||||
Db 10 aaugaaaaaaaaaaaaaa 27

RESULT 127
A08140
ID A08140 standard; DNA; 65 BP.
XX
AC A08140;
XX
DT 27-JUN-2000 (first entry)
XX
DE MOMP containing fusion protein construction PCR primer SEQ ID NO:27.
XX
KW Chlamydia trachomatis; fusion protein; major outer membrane protein;
KW MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;
KW infection; infectious disease; PCR primer; ss.
XX
OS Chlamydia trachomatis.
OS Synthetic.
XX
PN JP2000041678-A.
XX
PD 15-FEB-2000.
XX
PF 28-JUL-1998; 98JP-0213212.
XX
PR 28-JUL-1998; 98JP-0213212.
XX
PA (ELED) DENKI KAGAKU KOGYO KK.
XX

DR WPI: 2000-295780/26.
XX
PT A soluble fused protein useful for diagnosis of Chlamydia infection,
PT comprises at least part of major outer membrane protein (MOMP) of
PT Chlamydia trachomatis -
XX
PS Example 4; Page 32; 37pp; Japanese.
XX
CC The present invention describes fusion proteins (I) comprising at least
CC part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,
CC at least one hydrophilic polypeptide having no immunoreactivity to
CC human serum and their connected part. A08120 to A08125 encode
CC specifically claimed examples of the fusion proteins given in Y82388 to
CC Y82393. Also described is a method (A) for the detection of Chlamydia
CC trachomatis antibody using (I) as the antigen. (I) is used for the
CC diagnosis of Chlamydia trachomatis infectious diseases. The method can
CC diagnose Chlamydia trachomatis infectious diseases specifically in a
CC high sensitivity. A08126 to A08143 represent PCR oligonucleotide
CC primers used in the construction of the fusion proteins from the present
CC invention.
XX
SQ Sequence 65 BP; 40 A; 10 C; 2 G; 13 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 65;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3171 caataacaataacaataa 3188
||||||||||||||||||
Db 7 caataacaataacaataa 24

RESULT 128
A08141/c
ID A08141 standard; DNA; 65 BP.
XX
AC A08141;
XX
DT 27-JUN-2000 (first entry)
XX
DE MOMP containing fusion protein construction PCR primer SEQ ID NO:28.
XX
KW Chlamydia trachomatis; fusion protein; major outer membrane protein;
KW MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;
KW infection; infectious disease; PCR primer; ss.
XX
OS Chlamydia trachomatis.
OS Synthetic.
XX
PN JP2000041678-A.

XX
PD 15-FEB-2000.
XX
PF 28-JUL-1998; 98JP-0213212.
XX
PR 28-JUL-1998; 98JP-0213212.
XX
PA (ELED) DENKI KAGAKU KOGYO KK.
XX
DR WPI; 2000-295780/26.
XX
PT A soluble fused protein useful for diagnosis of Chlamydia infection,
PT comprises at least part of major outer membrane protein (MOMP) of
PT Chlamydia trachomatis -
XX
PS Example 4; Page 32; 37pp; Japanese.
XX
CC The present invention describes fusion proteins (I) comprising at least
CC part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,
CC at least one hydrophilic polypeptide having no immunoreactivity to
CC human serum and their connected part. A08120 to A08125 encode
CC specifically claimed examples of the fusion proteins given in Y82388 to
CC Y82393. Also described is a method (A) for the detection of Chlamydia
CC trachomatis antibody using (I) as the antigen. (I) is used for the
CC diagnosis of Chlamydia trachomatis infectious diseases. The method can
CC diagnose Chlamydia trachomatis infectious diseases specifically in a
CC high sensitivity. A08126 to A08143 represent PCR oligonucleotide
CC primers used in the construction of the fusion proteins from the present
CC invention.
XX
SQ Sequence 65 BP; 13 A; 2 C; 10 G; 40 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 65;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3171 caataacaataacaataa 3188
||||||||||||||||||
Db 59 CAATAACAATAACAATAA 42

RESULT 129
V90176
ID V90176 standard; cDNA; 115 BP.
XX
AC V90176;
XX
DT 15-FEB-1999 (first entry)
XX
DE EST clone DP776.

XX
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO9845436-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US06955.
XX
PR 10-APR-1997; 97US-0838821.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-070077/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
PS Claim 1; Page 456; 618pp; English.
XX
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 115 BP; 43 A; 21 C; 30 G; 21 T; 0 other;

Query Match 0.3%; Score 18; DB 20; Length 115;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5647 aaatccaaagcctctaag 5664
||||||||||||||||||

```
Db     56 aaatccaaagcctctaag 73


RESULT 130
Q37967
ID   Q37967 standard; cDNA; 226 BP.
XX
AC   Q37967;
XX
DT   30-JUN-1993  (first entry)
XX
DE   Sequence encoding purine-rich repeat (GA repeat) binding protien
DE   (GABP) subunit beta 2 from the point at which it diverges from
DE   beta 1.
XX
KW   GA binding protein; cis-regulatory element;
KW   VP16 mediated induction; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..48
FT                   /*tag=  a
XX
PN   WO9304166-A.
XX
PD   04-MAR-1993.
XX
PF   17-AUG-1992;   92WO-US06748.
XX
PR   16-AUG-1991;   91US-0746032.
XX
PA   (CARN-) CARNEGIE INST WASHINGTON.
XX
PI   Lamarco KL,  Mc Knight SL,  Thompson CC;
XX
DR   WPI; 1993-093998/11.
DR   P-PSDB; R33367.
XX
PT   DNA encoding GA binding protein sub-unit - allows investigation
PT   of sub-unit sequence motif functions, for control of rapid cell
PT   division e.g. in cancer
XX
PS   Claim 7; Fig 2B4; 68pp; English.
XX
CC   A cis-regulatory element required for virion associated protein VP16
CC   mediated induction of herpes simplex virus 1 (HSV1) immediate early
CC   (IE) genes consists of three imperfect repeats of the purine-rich
CC   hexanucleotide 5'-CGGAAR-3'. A protein complex capable of avid
CC   interaction with the purine-rich repeats (GA repeats) has been
CC   identified in soluble preparations of rat liver nucleic. This GA
```



```
CC   binding protein (GABP) consists of two separable subunits.
CC   Applicants have isolated cDNA clones encoding both subunits of GABP
CC   and have revealed that one (GABP alpha) is related to the Ets
CC   transforming protein, while the other (GABP beta) contains a
CC   series of 33-amino acid repeats related in sequence to a variety
CC   of proteins. The sequences for GABP-beta-1 and GABP-beta-2 are
CC   identical up to nucleotide 1130 except for a three nucleotide
CC   insertion (GTA) at posn. 828 of GABP-beta-1.
XX
SQ   Sequence 226 BP; 82 A; 27 C; 32 G; 85 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 14;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     205 aatgaaaaaaaaaaaaaa 222


RESULT 131
V60636
ID   V60636 standard; cDNA; 226 BP.
XX
AC   V60636;
XX
DT   23-NOV-1998  (first entry)
XX
DE   5'RACE product from paralysis tick HT-I cDNA.
XX
KW   Neurotoxin; HT-I; Australian paralysis tick; Ixodes holocyclus; PCR;
KW   primer; RACE; rapid amplification of cDNA ends; tryptic digest;
KW   immune response; cow; horse; goat; cat; dog; ss.
XX
OS   Synthetic.
OS   Ixodes holocyclus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..60
FT                   /*tag=  a
FT                   /note= "C-terminus of HT-I protein"
XX
PN   WO9747649-A1.
XX
PD   18-DEC-1997.
XX
PF   11-JUN-1997;   97WO-AU00366.
XX
PR   11-JUN-1996;   96AU-0000395.
XX
```



```
PA   (INSE-) INSEARCH LTD.
XX
PI   Broady KW,  Masina S,  Thurn MJ;
XX
DR   WPI; 1998-052243/05.
DR   P-PSDB; W68452.
XX
PT   Isolated nucleic acid encoding paralysis tick neurotoxin - used to
PT   immunise animals against the neurotoxin, specifically that of Ixodes
PT   holocyclus
XX
PS   Examples; Fig 11; 45pp; English.
XX
CC   This sequence represents the product generated by a 5'-RACE reaction
CC   using primer V60634, which based on the sequence of PCR fragment (V60633)
CC   of the neurotoxin HT-I gene from the Australian paralysis tick - Ixodes
CC   holocyclus.  When sequenced, the product revealed some differences
CC   between the amino acid sequence encoded by the cDNA and the sequence
CC   derived by amino acid sequencing the tryptic digest fragments. The new
CC   sequence was used to design new primers for obtaining the complete HT-I
CC   gene.  The polypeptide or nucleic acid encoding it can be used to raise
CC   a protective immune response against paralytic tick neurotoxin,
CC   especially in cows, horses, goats, cats and dogs.
XX
SQ   Sequence 226 BP; 105 A; 32 C; 48 G; 41 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 19;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     196 aatgaaaaaaaaaaaaaa 213


RESULT 132
V88786
ID   V88786 standard; cDNA; 256 BP.
XX
AC   V88786;
XX
DT   12-FEB-1999  (first entry)
XX
DE   EST clone HO332.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
```



```
OS   Homo sapiens.
XX
PN   WO9845437-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06956.
XX
PR   10-APR-1997;   97US-0837312.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070078/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 512; 641pp; English.
XX
CC   The present sequence represents an expressed sequence tag (EST), and is
CC   a polynucleotide of the invention. The polynucleotides of the invention
CC   are all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 256 BP; 80 A; 54 C; 55 G; 67 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      52 tctttgaaaatctgtcct 69
           ||||||||||||||||||
Db     130 tctttgaaaatctgtcct 147


RESULT 133
```

```
V89295/c
ID   V89295 standard; cDNA; 281 BP.
XX
AC   V89295;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone CH224.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 172; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 281 BP; 80 A; 53 C; 54 G; 94 T; 0 other;
```



```
  Query Match             0.3%;  Score 18;  DB 20;  Length 281;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db     77 AATGAAAAAAAAAAAAAA 60


RESULT 134
Q39730
ID   Q39730 standard; DNA; 355 BP.
XX
AC   Q39730;
XX
DT   20-MAY-1993  (first entry)
XX
DE   Expressed Sequence Tag human gene marker EST00365.
XX
KW   expressed sequence tag; human genome project; chromosome;
KW   human gene sequencing; PCR mapping; somatic cell hybrids;
KW   sublocalisation; gene tagging; tissue typing.
XX
OS   Synthetic.
XX
PN   WO9300353-A.
XX
PD   07-JAN-1993.
XX
PF   19-JUN-1992;   92WO-US05222.
XX
PR   20-JUN-1991;   91US-0716831.
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Venter JC;
XX
DR   WPI; 1993-036325/04.
XX
PT   Particular expressed sequence tags from human CDNA - corresponds
PT   to transcription prods. of genes, useful for tagging genes,
PT   mapping chromosomes and tissue typing
XX
PS   Claim 3; Page 117; 199pp; English.
XX
CC   This sequence represents an EST (expressed sequence tag) ESTs are markers
CC   for human genes actually transcribed in vivo. Unlike the random genomic
CC   DNA sequence tagged sites (STSs), ESTs point directly to expressed genes.
```



```
CC   The use of ESTs could facilitate the tagging of most expressed human
CC   genes within a few years at a fraction of the cost of complete genomic
CC   sequencing. Using PCR primers Q39419-Q39580 (sequences designed
CC   from the ESTs) sublocalisation of an EST can be achieved with panels of
CC   fragments from specific chromosomes or pools of large genomic clones in
CC   an analogous manner. This sequence represents EST00365.
XX
SQ   Sequence 355 BP; 87 A; 99 C; 61 G; 107 T; 1 other;

  Query Match             0.3%;  Score 18;  DB 14;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3173 ataacaataacaataatg 3190
          ||||||||||||||||||
Db    302 ataacaataacaataatg 319


RESULT 135
Q59142
ID   Q59142 standard; cDNA; 355 BP.
XX
AC   Q59142;
XX
DT   16-MAR-1994  (first entry)
XX
DE   Human brain Expressed Sequence Tag EST00365.
XX
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS   Homo sapiens.
XX
PN   WO9316178-A.
XX
PD   19-AUG-1993.
XX
PF   12-FEB-1993;   93WO-US01294.
XX
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Moreno RF,  Venter CJ;
XX
DR   WPI; 1993-272882/34.
XX
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
```



```
XX
PS   Example 1; Page 127; 500pp; English.
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00365 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM and hybridises to chromosome 1.
CC   See also Q59041-Q61440.
XX
SQ   Sequence 355 BP; 87 A; 99 C; 61 G; 107 T; 1 other;

  Query Match             0.3%;  Score 18;  DB 14;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3173 ataacaataacaataatg 3190
          ||||||||||||||||||
Db    302 ataacaataacaataatg 319


RESULT 136
Q10253
ID   Q10253 standard; DNA; 360 BP.
XX
AC   Q10253;
XX
DT   27-MAR-1991  (first entry)
XX
DE   Synthetic version of Alcohol Dehydrogenase 2 promoter.
XX
KW   urate oxidase; uricase; human growth hormone; ADH2; ds.
XX
OS   Synthetic.
XX
PN   EP408461-A.
XX
PD   16-JAN-1991.
XX
PF   13-JUL-1990;   90EP-0402023.
XX
PR   06-FEB-1990;   90FR-0001368.
PR   13-JUL-1989;   89FR-0009550.
PR   29-DEC-1989;   89FR-0017466.
XX
PA   (SNPI ) SANOFI SA.
```

```
XX
PI   Caput D,  Ferrara P,  Guillemot JC,  Kaghad M,  Legoux R;
PI   Loison G,  Labree,  Lupker J;
XX
DR   WPI; 1991-016644/03.
XX
PT   New protein with high urate oxidase activity - and recombinant
PT   DNA encoding it, vectors and transformed cells, used for treating
PT   hyperuricaemia, etc.
XX
PS   Example; Page 27; 68pp; French.
XX
CC   This is a synthetic fragment used in the construction of
CC   expression vectors for expression of a cDNA copy
CC   of the Aspergillus flavus urate oxidase gene. The ADH2 promoter
CC   sequence differs from the natural version (RUSSELL et al.,1983,
CC   J.Biol Chem. 256, 2674-2682) where several base pairs have been
CC   changed to introduce restriction sites.
CC   See also Q10245-Q10251 and Q10252 and Q10255.
XX
SQ   Sequence 360 BP; 123 A; 69 C; 68 G; 100 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 12;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db     79 aatgaaaaaaaaaaaaaa 96


RESULT 137
Z24905
ID   Z24905 standard; DNA; 367 BP.
XX
AC   Z24905;
XX
DT   02-DEC-1999  (first entry)
XX
DE   Human secreted protein gene 95 clone HCUHK65.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
```



```
OS   Homo sapiens.
XX
PN   WO9947540-A1.
XX
PD   23-SEP-1999.
XX
PF   18-MAR-1999;   99WO-US05804.
XX
PR   19-MAR-1998;   98US-0078563.
PR   19-MAR-1998;   98US-0078566.
PR   19-MAR-1998;   98US-0078573.
PR   19-MAR-1998;   98US-0078574.
PR   19-MAR-1998;   98US-0078576.
PR   19-MAR-1998;   98US-0078577.
PR   19-MAR-1998;   98US-0078578.
PR   19-MAR-1998;   98US-0078579.
PR   19-MAR-1998;   98US-0078581.
PR   01-APR-1998;   98US-0080312.
PR   01-APR-1998;   98US-0080313.
PR   01-APR-1998;   98US-0080314.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Rosen CA,  Yu G,  Young PE,  Feng P,  Soppet DR;
PI   Wei Y,  Endress GA,  Duan RD,  Kyaw H,  Ebner R,  Lafleur DW;
PI   Olsen HS,  Shi Y,  Moore PA;
XX
DR   WPI; 1999-562050/47.
DR   P-PSDB; Y41402.
XX
PT   New isolated human genes, useful for diagnosis and treatment of e.g.
PT   cancers, neurological disorders, immune diseases, inflammation or blood
PT   disorders -
XX
PS   Claim 1; Page 355; 484pp; English.
XX
CC   This sequence represents a nucleic acid molecule which encodes a secreted
CC   human protein.  The gene number, and the clone it is derived from, are
CC   detailed in the descriptor line.  The gene can be used to generate fusion
CC   proteins by linking to the gene to a human immunoglobulin Fc portion
CC   (e.g. Z24802) for increasing the stability of the fused protein as
CC   compared to the human protein only.
CC   The invention relates to 95 novel genes and their fragments (nucleic
CC   acid sequences: Z24811-Z24907; amino acid sequences Y41308-Y41404) which
CC   are useful for preventing, treating or ameliorating  medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 95 polynucleotides, based on
CC   which tissues they are most highly expressed in (see Z24811 for described
CC   uses).
```



```
XX
SQ   Sequence 367 BP; 139 A; 72 C; 70 G; 86 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    346 aatgaaaaaaaaaaaaaa 363


RESULT 138
Z54442/c
ID   Z54442 standard; DNA; 369 BP.
XX
AC   Z54442;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 925 partial DNA sequence SEQ ID NO:2833.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
```



```
DR   WPI; 2000-062150/05.
DR   P-PSDB; Y75680.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 7; Page 1332; 1453pp; English.
XX
CC   Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC   novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC   polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR
CC   primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 369 BP; 123 A; 77 C; 99 G; 70 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 21;  Length 369;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    224 cgttgcatttgatttctt 241
          ||||||||||||||||||
Db    346 CGTTGCATTTGATTTCTT 329


RESULT 139
X30899/c
ID   X30899 standard; DNA; 425 BP.
XX
AC   X30899;
XX
DT   20-MAY-1999  (first entry)
XX
DE   Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:176.
XX
KW   Streptococcus pneumoniae strain 0100993; vaccine; immune response;
KW   streptococcal infection; pneumococcal; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9737026-A1.
XX
```

```
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US05306.
XX
PR   22-AUG-1996;   96US-0025788.
PR   02-APR-1996;   96US-0014690.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1997-503111/46.
DR   P-PSDB; Y11317.
XX
PT   Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT   vaccines, drug screening, etc
XX
PS   Claim 5; Page 160; 354pp; English.
XX
CC   X30724 to X30946 represent genomic DNA sequences isolated from
CC   Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC   encode the novel proteins given in Y11114 to Y11367. The proteins,
CC   isolated from Streptococcus pneumoniae, can be used in vaccines against
CC   streptococcal infections and in assays for identifying compounds that
CC   inhibit or activate the activity of the proteins. The antagonists can
CC   be used to treat an individual having need to inhibit a bacterial
CC   protein. Vectors expressing the proteins can be used to induce a
CC   protective immune response in mammals.
XX
SQ   Sequence 425 BP; 154 A; 81 C; 86 G; 104 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 18;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4744 tttcttttcttgctagatt 4761
        ||||||||||||||||||
Db     90 TTTCTTTCTTGCTAGATT 73


RESULT 140
X97675
ID   X97675 standard; DNA; 471 BP.
XX
AC   X97675;
XX
DT   13-SEP-1999  (first entry)
XX
```



```
DE   Extended human secreted protein coding sequence, SEQ ID NO. 139.
XX
KW   Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW   cellular differentiation; immune system regulator; anti-inflammatory;
KW   haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW   reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW   genetic disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9931236-A2.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-IB02122.
XX
PR   10-AUG-1998;   98US-0096116.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-385906/32.
DR   P-PSDB; Y35991.
XX
PT   New isolated human secreted proteins
XX
PS   Claim 1; Page 267-268; 516pp; English.
XX
CC   This sequence represents an extended human secreted protein coding
CC   sequence of the invention. The secreted proteins can be used in treating
CC   or controlling a variety of human conditions. The secreted proteins may
CC   act as cytokines or may affect cellular proliferation or differentiation
CC   or may act as immune system regulators, haematopoiesis regulators, tissue
CC   growth regulators, regulators of reproductive hormones or cell movement
CC   or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC   tumour inhibition activity. The DNAs can be used in forensic procedures
CC   to identify individuals or in diagnostic procedures to identify
CC   individuals having genetic diseases resulting from abnormal expression of
CC   the genes corresponding to the extended cDNAs. They are also useful for
CC   constructing a high resolution map of the human chromosomes. They can
CC   also be used for gene therapy to control or treat genetic diseases.
XX
SQ   Sequence 471 BP; 105 A; 120 C; 120 G; 126 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 471;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
```



```
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    454 aatgaaaaaaaaaaaaaa 471


RESULT 141
V86280/c
ID   V86280 standard; cDNA; 491 BP.
XX
AC   V86280;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone AJ143.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 182; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
```



```
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 491 BP; 104 A; 137 C; 111 G; 138 T; 1 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 491;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3961 ggtttccaaaaatcttaa 3978
          ||||||||||||||||||
Db    250 GGTTTCCAAAAATCTTAA 233


RESULT 142
V09106
ID   V09106 standard; cDNA; 507 BP.
XX
AC   V09106;
XX
DT   28-MAY-1998  (first entry)
XX
DE   3' nucleotide sequence of clone D305_2.
XX
KW   D305_2 clone; cDNA library; cell proliferation;  cell differentiation;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   polyA_signal    474..480
FT                   /*tag= a
XX
PN   US5708157-A.
XX
PD   13-JAN-1998.
XX
PF   26-JUL-1996;   96US-0686878.
XX
PR   26-JUL-1996;   96US-0686878.
XX
PA   (GEMY ) GENETICS INST INC.
PA   (JACO/) JACOBS K.
XX
```

```
PI   Evans C,  Jacobs K,  LaVallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1998-109352/10.
XX
PT   Human testes derived cDNA and protein(s) - may be useful for
PT   laboratory use, or for stimulating cell proliferation and
PT   differentiation
XX
PS   Disclosure; Column 61-62; 48pp; English.
XX
CC   This sequence is the 3' nucleotide sequence for the D305_2 clone, which
CC   is isolated from a human PBMC cDNA library.  When the D305_2 sequence
CC   was searched in the Genbank database, the clone demonstrated homology
CC   with an EST identified as "yi14g05.s1 Homo sapiens cDNA clone 139526 3'".
XX
SQ   Sequence 507 BP; 165 A; 70 C; 83 G; 189 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 19;  Length 507;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     490 aatgaaaaaaaaaaaaaa 507


RESULT 143
X77537
ID   X77537 standard; cDNA; 507 BP.
XX
AC   X77537;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human secreted protein D305_2 cDNA fragment 3.
XX
KW   Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW   bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW   cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW   haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW   cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
OS   Homo sapiens.
XX
PN   WO9926972-A1.
XX
PD   03-JUN-1999.
XX
```



```
PF   17-NOV-1998;   98WO-US24614.
XX
PR   20-OCT-1998;   98US-0175928.
PR   21-NOV-1997;   97US-0976110.
PR   18-MAY-1998;   98US-0080478.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Collins-Racie LA,  Evans C,  Jacobs K,  Lavallie ER;
PI   McCoy JM,  Merberg D,  Treacy M;
XX
DR   WPI; 1999-357813/30.
DR   P-PSDB; Y08629.
XX
PT   New polynucleotides encoding secreted proteins
XX
PS   Claim 36; Page 115; 142pp; English.
XX
CC   This invention describes novel human secreted proteins encoded by
CC   polynucleotides isolated from human adult testes, adult brain, adult
CC   blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC   libraries. The products of the invention are predicted to have biological
CC   activities which would make them suitable for treating, preventing or
CC   ameliorating medical conditions in humans and animals, although no
CC   supporting data is given. Suggested activities include nutritional
CC   activity, cytokine and cell proliferation/differentiation activity,
CC   immune stimulating (e.g. as vaccines) or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, and tumour
CC   inhibition activity. The polynucleotides are also stated to be useful
CC   for gene therapy.
XX
SQ   Sequence 507 BP; 165 A; 70 C; 83 G; 189 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 507;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     490 aatgaaaaaaaaaaaaaa 507


RESULT 144
Z59479
ID   Z59479 standard; cDNA; 507 BP.
XX
AC   Z59479;
```



```
XX
DT   11-APR-2000  (first entry)
XX
DE   Human secreted protein D305_2 3'portion polynucleotide sequence.
XX
KW   Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW   placental pathology; metastasis inhibition; nutritional activity;
KW   immune stimulator; haematopoiesis regulator; tissue growth;
KW   tumour inhibitor; anti-inflammatory; clone D305_2; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9960020-A1.
XX
PD   25-NOV-1999.
XX
PF   17-MAY-1999;   99WO-US10915.
XX
PR   18-MAY-1998;   98US-0080478.
PR   20-OCT-1998;   98US-0175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Mi S,  Treacy M;
XX
DR   WPI; 2000-116311/10.
XX
PT   New polynucleotides encoding secreted cDNA libraries, used to develop
PT   products for the diagnosis and treatment of neoplastic disease
XX
PS   Disclosure; Page 122; 149pp; English.
XX
CC   This is the human secreted protein D305_2 3' portion nucleotide
CC   sequence, obtained from a human adult blood cDNA library. The invention
CC   relates to secreted human and murine proteins. The polynucleotides and
CC   proteins are predicted to have biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical conditions
CC   in humans and animals. Detection of the levels of the proteins can be
CC   used for the diagnosis of e.g. pre-eclampsia, placental pathology or
CC   cancer. Agents which modulate the expression or function of the proteins
CC   may be used for treating a neoplastic disease and inhibiting metastasis.
CC   Other suggested activities include nutritional activity (e.g. in feeds),
CC   cytokine and cell proliferation/differentiation activity, immune
CC   stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis
CC   regulating activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC   invasion suppressor activity, and tumour inhibition activity. The
CC   polynucleotide sequences are also stated to be useful for gene therapy.
XX
```



```
SQ   Sequence 507 BP; 165 A; 70 C; 83 G; 189 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 21;  Length 507;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     490 aatgaaaaaaaaaaaaaa 507


RESULT 145
Z54438/c
ID   Z54438 standard; DNA; 516 BP.
XX
AC   Z54438;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 925 partial DNA sequence SEQ ID NO:2825.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
```

DR P-PSDB: Y75676.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 1331; 1453pp; English.
XX
CC Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR
CC primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 516 BP; 158 A; 107 C; 153 G; 98 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 516;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 cgttgcatttgatttctt 241
||||||||||||||||||
Db 493 CGTTGCATTTGATTTCTT 476

RESULT 146
Z54440/c
ID Z54440 standard; DNA; 516 BP.
XX
AC Z54440;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrheae ORF 925 partial DNA sequence SEQ ID NO:2829.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria gonorrheae.
XX
PN WO9957280-A2.
XX

PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; Y75678.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 1331; 1453pp; English.
XX
CC Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR
CC primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 516 BP; 158 A; 107 C; 153 G; 98 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 516;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 cgttgcatttgatttctt 241
||||||||||||||||||

Db 493 CGTTGCATTTGATTTCTT 476

RESULT 147
T98751/c
ID T98751 standard; DNA; 518 BP.
XX
AC T98751;
XX
DT 10-NOV-1998 (first entry)
XX
DE DNA encoding a S. pneumoniae protein of unknown function.
XX
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis; ss.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 177..437
FT /*tag= a
XX
PN WO9743303-A1.
XX
PD 20-NOV-1997.
XX
PF 14-MAY-1997; 97WO-US07950.
XX
PR 14-MAY-1996; 96US-0017670.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;
XX
DR WPI; 1998-008793/01.
DR P-PSDB; W38709.
XX
PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
XX
PS Claim 4; Page 238; 483pp; English.
XX
CC This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function, and represents a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain

CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.
XX
SQ Sequence 518 BP; 185 A; 98 C; 110 G; 125 T; 0 other;

Query Match 0.3%; Score 18; DB 19; Length 518;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4744 tttctttcttgctagatt 4761
||||||||||||||||||
Db 90 TTTCTTTCTTGCTAGATT 73

RESULT 148
Z54441/c
ID Z54441 standard; DNA; 522 BP.
XX
AC Z54441;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 925 partial DNA sequence SEQ ID NO:2831.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; Y75679.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
PS Claim 7; Page 1332; 1453pp; English.
XX
CC Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR
CC primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 522 BP; 163 A; 108 C; 152 G; 99 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 522;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 cgttgcatttgatttctt 241
||||||||||||||||||
Db 496 CGTTGCATTTGATTTCTT 479

RESULT 149
X97614

ID X97614 standard; DNA; 542 BP.
XX
AC X97614;
XX
DT 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein coding sequence, SEQ ID NO. 78.
XX
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease; ss.
XX
OS Homo sapiens.
XX
PN WO9931236-A2.
XX
PD 24-JUN-1999.
XX
PF 17-DEC-1998; 98WO-IB02122.
XX
PR 10-AUG-1998; 98US-0096116.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
DR WPI; 1999-385906/32.
DR P-PSDB; Y35930.
XX
PT New isolated human secreted proteins
XX
PS Claim 1; Page 212; 516pp; English.
XX
CC This sequence represents an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.

XX
SQ Sequence 542 BP; 132 A; 146 C; 126 G; 138 T; 0 other;

Query Match 0.3%; Score 18; DB 20; Length 542;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 518 aatgaaaaaaaaaaaaaa 535

RESULT 150
X20418
ID X20418 standard; DNA; 545 BP.
XX
AC X20418;
XX
DT 04-MAY-1999 (first entry)
XX
DE Human secreted protein gene 7.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9906423-A1.
XX
PD 11-FEB-1999.
XX
PF 29-JUL-1998; 98WO-US15949.
XX
PR 19-AUG-1997; 97US-0056730.
PR 30-JUL-1997; 97US-0054209.
PR 30-JUL-1997; 97US-0054211.
PR 30-JUL-1997; 97US-0054212.
PR 30-JUL-1997; 97US-0054213.
PR 30-JUL-1997; 97US-0054214.
PR 30-JUL-1997; 97US-0054215.
PR 30-JUL-1997; 97US-0054217.
PR 30-JUL-1997; 97US-0054218.
PR 30-JUL-1997; 97US-0054234.
PR 30-JUL-1997; 97US-0054236.

PR 18-AUG-1997; 97US-0055968.
PR 18-AUG-1997; 97US-0055969.
PR 18-AUG-1997; 97US-0055972.
PR 19-AUG-1997; 97US-0056534.
PR 19-AUG-1997; 97US-0056543.
PR 19-AUG-1997; 97US-0056554.
PR 19-AUG-1997; 97US-0056561.
PR 19-AUG-1997; 97US-0056727.
PR 19-AUG-1997; 97US-0056729.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Endress GA, Fan P, Feng P, Kyaw H, Lafleur DW;
PI Li Y, Moore PA, Rosen CA, Ruben SM, Shi Y, Wei Y;
PI Zeng Z;
XX
DR WPI; 1999-153691/13.
DR P-PSDB; Y00264.
XX
PT New isolated human genes and the secreted polypeptides they encode
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 214; 312pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number is given in the descriptor line. The gene
CC can be used to generate fusion proteins by linking to the gene to a human
CC immunoglobulin Fc portion (e.g. X20403) for increasing the stability of
CC the fused protein as compared to the human protein only.
CC The invention relates to 83 novel genes and their fragments (nucleic acid
CC sequences: X20412-X20499; amino acid sequences Y00258-Y00377) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 86 polynucleotides, based on
CC which tissues they are most highly expressed in (see X20412 for described
CC uses).
XX
SQ Sequence 545 BP; 137 A; 143 C; 164 G; 97 T; 4 other;

Query Match 0.3%; Score 18; DB 20; Length 545;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 503 aatgaaaaaaaaaaaaaa 520

```
RESULT 151
V02880
ID   V02880 standard; cDNA; 566 BP.
XX
AC   V02880;
XX
DT   11-MAY-1998  (first entry)
XX
DE   Human HMGI-C aberrant form 1.
XX
KW   High mobility group protein; HMGI-C; NAG; human; treatment; modulator;
KW   multiple tumour aberration growth gene; vascular development;
KW   angiogenesis; vascularisation; endometriosis; contraception
KW   tissue regeneration; ss.
XX
OS   Homo sapiens.
XX
PN   DE19548122-A1.
XX
PD   26-JUN-1997.
XX
PF   21-DEC-1995;   95DE-1048122.
XX
PR   21-DEC-1995;   95DE-1048122.
XX
PA   (BULL/) BULLERDIEK J.
XX
PI   Bullerdiek J;
XX
DR   WPI; 1997-333837/31.
XX
PT   DNA sequences representing aberrant forms of human high mobility
PT   group protein genes - useful for treatment of endometriosis and
PT   tumours, or for modulating vascularisation, etc
XX
PS   Claim 1; Fig 1; 58pp; German.
XX
CC   V02880-V02898 are cDNA sequences that encode aberrant forms of the human
CC   high mobility group protein (HMG) gene, HMGI-C, which is located on
CC   chromosome 12. These sequences encode the DNA binding part of the
CC   translation product but not the protein binding domain. These proteins,
CC   antibodies derived from these proteins or expression modulators of the
CC   HMGI-C protein can be used in kits to modulate vascular development. Such
CC   kits can reduce, block or stimulate angiogenesis or vascularisation and
CC   can improve vascular provision in myocardium damaged by infarction. Such
CC   proteins can also be used to treat endometriosis and tumours, for
CC   contraception (local or oral) and for tissue regeneration, especially in
CC   degenerating or damaged tissue. The regeneration method can be applied to
CC   tissues which are currently impossible or difficult to regenerate and the
CC   use of biological material with attendant risks of viral transmission and
```



```
CC   anaphylactic shock, is avoided.
XX
SQ   Sequence 566 BP; 186 A; 136 C; 140 G; 104 T; 0 other;

  Query Match             0.3%;  Score 18;  DB 18;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    544 aatgaaaaaaaaaaaaaa 561


RESULT 152
V03044
ID   V03044 standard; DNA; 613 BP.
XX
AC   V03044;
XX
DT   18-JUN-1998  (first entry)
XX
DE   Rat EX clone nucleotide sequence.
XX
KW   Angiotensin IV; neuroactive; Alzheimer's disease; modulator; dementia;
KW   neuronal development; vasoeffective; neuropathy; brain injury;
KW   LVV-haemorphin-7; beta-globin; sheep; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             23..466
FT                   /*tag= a
XX
PN   WO9801465-A1.
XX
PD   15-JAN-1998.
XX
PF   09-JUL-1997;   97WO-AU00436.
XX
PR   09-JUL-1996;   96AU-0000893.
XX
PA   (FLOR-) FLOREY INST EXPERIMENTAL PHYSIOLOGY.
XX
PI   Aldred PG,  Chai SY,  Lew RA,  Mendelsohn FAO,  Moeller I;
PI   Smith IA;
XX
DR   WPI; 1998-100995/09.
DR   P-PSDB; W37248.
XX
PT   Neuro-active peptide with activities of angiotensin IV and related
```



```
PT   DNA - antagonists and oligo:nucleotide(s), useful for modulating
PT   neuronal development and activity of motor and cholinergic neurons
XX
PS   Example 7; Fig 13; 65pp; English.
XX
CC   This is the nucleotide sequence of a rat EX clone. This encodes a
CC   precursor of a neuroactive peptide LVV-haemorphin-7. Specific primers
CC   corresponding to the region of the sheep beta-globin gene encoding the
CC   LVV-haemorphin-7 sequence are used for screening and cloning beta-globin
CC   like sequences from rat cDNA library by PCR amplification. The
CC   neuroactive peptide is an agonist of the angiotensin IV receptor and is
CC   its endogenous ligand in the brain. It can modulate learning or
CC   behaviour and has vasoactive effects and can dilate cerebral arteries,
CC   increase renal blood flow and stereotypy behaviour, facilitate memory
CC   retrieval, induce neurite remodelling, and alleviate effects of spinal
CC   cord injuries. This peptide and its antagonists are used to modulate
CC   activity of motor and cholinergic neurons and neuronal development.
CC   Specific applications are in treatment of dementia, Alzheimer's disease,
CC   neurodegeneration (involving cholinergic, motor or sensory pathways),
CC   sensory and motor peripheral neuropathy, brain injury and spinal cord
CC   injury (caused by one or more of trauma, hypoxia and/or vascular
CC   disease). Antibodies specific for the peptide can be used to determine
CC   its amount of in tissues and tissue regions.
XX
SQ   Sequence 613 BP; 159 A; 147 C; 155 G; 152 T; 0 other;

  Query Match             0.3%;  Score 18;  DB 19;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     53 ctttgaaaatctgtccta 70
          ||||||||||||||||||
Db    551 ctttgaaaatctgtccta 568


RESULT 153
V86929/c
ID   V86929 standard; cDNA; 634 BP.
XX
AC   V86929;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone BK299.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
```



```
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 395; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 634 BP; 200 A; 114 C; 121 G; 198 T; 1 other;

  Query Match             0.3%;  Score 18;  DB 20;  Length 634;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4346 tctttaaagtctgcatgg 4363
          ||||||||||||||||||
Db    504 TCTTTAAAGTCTGCATGG 487


RESULT 154
```

```
X22221
ID   X22221 standard; DNA; 637 BP.
XX
AC   X22221;
XX
DT   18-MAY-1999  (first entry)
XX
DE   Human secreted protein gene 11 clone HWTBK81.
XX
KW   Human; secreted protein; gene therapy; protein therapy; cancer; weight;
KW   tumour; chromosome mapping; forensic; haematological disease; allergy;
KW   inflammation; cell proliferation; viral infection; wound healing;
KW   modulation; appetite; behaviour; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO9903990-A1.
XX
PD   28-JAN-1999.
XX
PF   15-JUL-1998;   98WO-US14613.
XX
PR   18-AUG-1997;   97US-0056361.
PR   16-JUL-1997;   97US-0052661.
PR   16-JUL-1997;   97US-0052870.
PR   16-JUL-1997;   97US-0052871.
PR   16-JUL-1997;   97US-0052872.
PR   16-JUL-1997;   97US-0052873.
PR   16-JUL-1997;   97US-0052874.
PR   16-JUL-1997;   97US-0052875.
PR   22-JUL-1997;   97US-0053440.
PR   22-JUL-1997;   97US-0053441.
PR   22-JUL-1997;   97US-0053442.
PR   18-AUG-1997;   97US-0055683.
PR   18-AUG-1997;   97US-0055724.
PR   18-AUG-1997;   97US-0055725.
PR   18-AUG-1997;   97US-0055726.
PR   18-AUG-1997;   97US-0055946.
PR   18-AUG-1997;   97US-0055952.
PR   18-AUG-1997;   97US-0055985.
PR   18-AUG-1997;   97US-0055989.
PR   18-AUG-1997;   97US-0056359.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Duan R,  Feng P,  Ferrie AM,  Florence KA,  Fouad J;
PI   Greene JM,  Hu J,  Ni J,  Rosen CA,  Ruben SM,  Young PE;
PI   Yu G;
XX
DR   WPI; 1999-132234/11.
DR   P-PSDB; Y01393.
```



```
XX
PT   New nucleic acids encoding secreted human proteins - potentially
PT   useful for treating and diagnosing diseases and identifying specific
PT   binding agents
XX
PS   Claim 4; Page 175; 251pp; English.
XX
CC   The invention relates to nucleic acid sequences (X22211 to X22282)
CC   encoding human secreted proteins (Y01383 to Y01454). The secreted protein
CC   gene sequences are deposited with the ATCC under deposit number ATCC
CC   209138, 209139 or 209141. Host cells containing vectors comprising the
CC   nucleic acid sequences are used for the recombinant expression of the
CC   secreted proteins. The polynucleotide and amino acid sequences are useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. Pathological conditions can be also diagnosed by
CC   determining the amount of the new polypeptides in a sample or by the
CC   presence of mutations in the new polynucleotides. The nucleic acid
CC   sequences, or its fragments, are useful for chromosome identification and
CC   mapping; as antisense and triplex-forming therapeutics; in gene therapy;
CC   for (forensic) identification of individuals; as molecular weight
CC   markers; to identify related sequences or specific mRNA; in preparation
CC   of oligomers and to raise anti-DNA antibodies. Antibodies are useful as
CC   immunoassay reagents (including for in vivo imaging) and therapeutically
CC   to inhibit or activate particular polypeptides. A very wide range of
CC   disorders may be treated with the polynucleotide and polypeptide
CC   sequences, e.g. autoimmune or haematological diseases, allergy,
CC   inflammation, cancer or other forms of cell proliferation, viral or other
CC   infections. The sequences may also be useful in wound healing, to
CC   modulate differentiation of embryonic stem cells, to modulate weight,
CC   appetite, behaviour etc. and as food additive or preservative. The
CC   present sequence represents a gene encoding a human secreted protein
CC   (see descriptor line for gene number and clone identification).
XX
SQ   Sequence 637 BP; 164 A; 164 C; 159 G; 149 T; 1 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 637;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    609 aatgaaaaaaaaaaaaaa 626


RESULT 155
Z96205/c
ID   Z96205 standard; DNA; 664 BP.
XX
AC   Z96205;
XX
```



```
DT   10-APR-2000  (first entry)
XX
DE   S. pneumoniae derived DNA from ORF #33.
XX
KW   Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW   bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9806734-A1.
XX
PD   19-FEB-1998.
XX
PF   15-AUG-1997;   97WO-US14436.
XX
PR   16-AUG-1996;   96US-0024022.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Lonetto MA,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1998-159452/14.
DR   P-PSDB; Y85827.
XX
PT   Streptococcus pneumoniae proteins and related DNA - useful for
PT   screening compounds for antibacterial activity
XX
PS   Claim 4; Page 74-75; 640pp; English.
XX
CC   This invention describes novel isolated Streptococcus pneumoniae
CC   polynucleotides (see Z96173-Z96494) and their encoded proteins (see
CC   Y85792-Y86182). The DNA, vectors and host cells described in the method
CC   of the invention are useful for the recombinant expression of the
CC   polypeptides. The polypeptides are useful for treatment or prevention of
CC   disease, or diagnosis of disease related to expression or activity of
CC   such a polypeptide. They can also be used to screen for compounds which
CC   interact with and inhibit or activate such a polypeptide. The
CC   polypeptides (or DNA encoding them, via gene therapy) are also useful
CC   for inducing an immunological response in a mammal. The antagonists are
CC   useful to inhibit such bacterial polypeptides. The polypeptides are
CC   particularly useful to identify antimicrobial compounds and antibiotics.
CC   They are also useful to determine their role in pathogenesis of
CC   infection, dysfunction and disease.
XX
SQ   Sequence 664 BP; 232 A; 129 C; 139 G; 164 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 19;  Length 664;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy   4744 tttctttcttgctagatt 4761
          ||||||||||||||||||
Db    148 TTTCTTTCTTGCTAGATT 131


RESULT 156
X97861
ID   X97861 standard; cDNA; 675 BP.
XX
AC   X97861;
XX
DT   23-SEP-1999  (first entry)
XX
DE   Human secreted protein encoding cDNA #49.
XX
KW   Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW   diagnostic; gene therapy; chromosome mapping; secretion vector; ss.
XX
OS   Homo sapiens.
XX
PN   WO9925825-A2.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-IB01862.
XX
PR   04-SEP-1998;   98US-0099273.
PR   13-NOV-1997;   97US-0066677.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-347472/29.
DR   P-PSDB; Y36177.
XX
PT   Extended cDNAs encoding secreted proteins
XX
PS   Claim 1; Page 251-252; 307pp; English.
XX
CC   X97813-X97906 represent extended cDNA's which encode novel human secreted
CC   proteins (see Y36129-Y36222) and which have cytostatic, thrombotic and
CC   osteopathic activity. The extended cDNAs can be used to express secreted
CC   proteins or parts of them or to obtain antibodies capable of binding to
CC   the secreted proteins. They may also be used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. Uses also include design
```

```
CC   of expression vectors and secretion vectors.
XX
SQ   Sequence 675 BP; 226 A; 97 C; 110 G; 242 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 675;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     658 aatgaaaaaaaaaaaaaa 675


RESULT 157
Z16920/c
ID   Z16920 standard; cDNA; 676 BP.
XX
AC   Z16920;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:4390.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones WL,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
```



```
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 2079; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in Z12532 to Z17779. Also described is a
CC   method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in Z12532 to Z17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 676 BP; 184 A; 129 C; 144 G; 209 T; 10 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 676;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      83 AATGAAAAAAAAAAAAAA 66


RESULT 158
A26408
ID   A26408 standard; cDNA; 680 BP.
XX
AC   A26408;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 63 SEQ ID NO:73.
XX
```



```
KW   Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW   antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW   osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW   antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW   immune disease; inflammation; blood disorder; tumour; ss.
XX
OS   Homo sapiens.
XX
PN   WO200006698-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-US17130.
XX
PR   30-JUL-1998;   98US-0094657.
PR   05-AUG-1998;   98US-0095486.
PR   06-AUG-1998;   98US-0095454.
PR   06-AUG-1998;   98US-0095455.
PR   12-AUG-1998;   98US-0096319.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Komatsoulis GA,  Rosen CA,  Ruben SM,  Duan R,  Moore PA,  Shi Y;
PI   Lafleur D,  Wei Y,  Ni J,  Florence KA,  Young PE,  Brewer LA;
PI   Soppet DR,  Endress GA,  Ebner R,  Olsen HS,  Mucenski M;
XX
DR   WPI; 2000-195282/17.
DR   P-PSDB; Y91513.
XX
PT   New isolated human genes and the secreted polypeptides they encode,
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 416; 634pp; English.
XX
CC   The polynucleotide sequences given in A26346 to A26458 encode the human
CC   secreted proteins given in Y91451 to Y91691. The human secreted proteins
CC   can have activities based on the tissues and cells they are expressed in.
CC   Examples of the activities are: cytostatic; immunosuppressive; antiHIV;
CC   antiinflammatory; nootropic; neuroprotective; antiallergic; osteopathic;
CC   antiarthritic; antibacterial; antidiabetic; antiasthma; antipsoriatic;
CC   and cardiant. The polynucleotides and their corresponding secreted
CC   proteins are useful for preventing, treating or ameliorating medical
CC   conditions, e.g. by protein or gene therapy. Also pathological conditions
CC   can be diagnosed by determining the amount of the proteins in a sample or
CC   by determining the presence of mutations in the polynucleotides. Specific
CC   uses are described for each of the polynucleotides, based on which
CC   tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, tumours,
CC   neurodegenerative disorders, developmental abnormalities and foetal
CC   deficiencies, blood disorders, diseases of the immune system,
```



```
CC   autoimmune diseases, hepatic and renal disease, inflammation,
CC   allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC   osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC   transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC   cardiovascular disorders, reproductive disorders, gastrointestinal
CC   disorders, respiratory disorders and metabolic disorders. The proteins
CC   or polynucleotides can also be used as food additives or preservatives.
CC   The proteins are also useful for identifying their binding partners.
CC   A26337 to A26345 and Y91450 are sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 680 BP; 163 A; 185 C; 167 G; 156 T; 9 other;


  Query Match             0.3%;  Score 18;  DB 21;  Length 680;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     635 aatgaaaaaaaaaaaaaa 652


RESULT 159
V24706
ID   V24706 standard; DNA; 693 BP.
XX
AC   V24706;
XX
DT   15-JUN-1998  (first entry)
XX
DE   H. pylori ORF 14eel1217orf3.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..693
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   WO9737044-A1.
XX
PD   09-OCT-1997.
XX
PF   27-MAR-1997;   97WO-US05223.
XX
PR   06-DEC-1996;   96US-0761318.
```

PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
PA (ASTR) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR WPI; 1997-503122/46.
DR P-PSDB; W55297.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Claims 5,6; Page 195; 1145pp; English.
XX
CC This sequence encodes a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SQ Sequence 693 BP; 221 A; 156 C; 134 G; 182 T; 0 other;

Query Match 0.3%; Score 18; DB 18; Length 693;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4528 ttttagagcaaattgatc 4545
||||||||||||||||||
Db 245 ttttagagcaaattgatc 262

RESULT 160
Z56708
ID Z56708 standard; cDNA; 718 BP.
XX
AC Z56708;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human transmembrane protein HTMPN-11 encoding cDNA.
XX
KW Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW antiproliferative; neuroprotective; immune disorder;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder; ss.
XX
OS Homo sapiens.
XX
PN WO9961471-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11904.
XX
PR 29-MAY-1998; 98US-0087260.
PR 02-JUL-1998; 98US-0091674.
PR 02-OCT-1998; 98US-0102954.
PR 24-NOV-1998; 98US-0109869.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
PI Au-Young J;
XX
DR WPI; 2000-072605/06.
DR P-PSDB; Y57887.
XX
PT Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,
PT neurological, gastrointestinal, developmental and cell proliferative
PT disorders -
XX
PS Claim 9; Page 184; 229pp; English.
XX
CC Z56698 to Z56776 encode Y57877 to Y57955 which represent human
CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
CC The transmembrane protein have immunospecific, antiproliferative and
CC neuroprotective activities. The human transmembrane proteins,
CC polynucleotides encoding them and other compositions and methods from
CC the present invention, can be used for the diagnosis, treatment or

CC prevention of immune, reproductive, smooth muscle, neurological,
CC gastrointestinal, developmental and cell proliferative disorders. The
CC HTMPN's can be used to treat or prevent disorders associated with a
CC decreased expression or activity of HTMPN.
XX
SQ Sequence 718 BP; 200 A; 130 C; 173 G; 215 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 718;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 698 aatgaaaaaaaaaaaaaa 715

RESULT 161
Z17597
ID Z17597 standard; cDNA; 741 BP.
XX
AC Z17597;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:5070.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 2411; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in Z12532 to Z17779. Also described is a
CC method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in Z12532 to Z17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 741 BP; 193 A; 137 C; 145 G; 253 T; 13 other;

Query Match 0.3%; Score 18; DB 20; Length 741;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 657 aatgaaaaaaaaaaaaaa 674

RESULT 162
Z15143
ID Z15143 standard; cDNA; 742 BP.
XX
AC Z15143;
XX
DT 12-OCT-1999 (first entry)

XX
DE Human gene expression product cDNA sequence SEQ ID NO:2612.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 1272; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in Z12532 to Z17779. Also described is a
CC method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in Z12532 to Z17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an

CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 742 BP; 158 A; 191 C; 196 G; 179 T; 18 other;

Query Match 0.3%; Score 18; DB 20; Length 742;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 682 aatgaaaaaaaaaaaaaa 699

RESULT 163
A37024
ID A37024 standard; cDNA; 756 BP.
XX
AC A37024;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1018 (UNQ501) cDNA sequence SEQ ID NO:7.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.

PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.

PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.

PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2000-237871/20.
DR P-PSDB; Y99342.
XX
PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS Claim 2; Fig 5; 773pp; English.
XX
CC A37022 to A37144 encode the new isolated human transmembrane, receptor
CC or secreted PRO polypeptides given in Y99340 to Y99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. A37145 to A37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 756 BP; 155 A; 226 C; 208 G; 167 T; 0 other;

Query Match 0.3%; Score 18; DB 21; Length 756;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 734 aatgaaaaaaaaaaaaaa 751

RESULT 164
V25011
ID V25011 standard; DNA; 774 BP.
XX

AC V25011;
XX
DT 07-JUL-1998 (first entry)
XX
DE H. pylori inner membrane protein ORF 11ap20714_5271967_c1_60.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 1..774
FT /*tag= a
XX
PN WO9737044-A1.
XX
PD 09-OCT-1997.
XX
PF 27-MAR-1997; 97WO-US05223.
XX
PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
PA (ASTR) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR WPI; 1997-503122/46.
DR P-PSDB; W55602.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Claims 5,6; Pages 388-389; 1145pp; English.
XX
CC This sequence encodes a H. pylori cell envelope inner membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of

CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 100 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SQ Sequence 774 BP; 243 A; 176 C; 154 G; 201 T; 0 other;

Query Match 0.3%; Score 18; DB 18; Length 774;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4528 ttttagagcaaattgatc 4545
||||||||||||||||||
Db 323 ttttagagcaaattgatc 340

RESULT 165
Z52877/c
ID Z52877 standard; cDNA; 802 BP.
XX
AC Z52877;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor cDNA library derived EST fragment #20.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment; ds.
XX
OS Homo sapiens.
XX
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (METAG-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-621386/54.

DR P-PSDB; Y73871, Y73872, Y73873.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
PS Claim 2; Page 198; 502pp; German.
XX
CC This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. Z52858-Z53014 represent
CC expressed sequence tag (EST) fragments derived from a human pancreatic
CC tumor cDNA library and which encode the proteins represented in
CC Y73814-Y74252.
XX
SQ Sequence 802 BP; 119 A; 216 C; 280 G; 187 T; 0 other;

Query Match 0.3%; Score 18; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 22 AATGAAAAAAAAAAAAAA 5

RESULT 166
X14350
ID X14350 standard; DNA; 810 BP.
XX
AC X14350;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 292 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 13..786
FT /*tag= a
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.

```
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   P-PSDB; W98631.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 1; Page 1286-1287; 2054pp; English.
XX
CC   This sequence represents a polynucleotide of the invention. It was
CC   isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC   The polypeptides can be used for preventing or treating Helicobacter
CC   infections, and gastroduodenal diseases associated with these
CC   infections, including acute, chronic, and atrophic gastritis, and peptic
CC   ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC   for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence 810 BP; 258 A; 179 C; 157 G; 215 T; 1 other;


  Query Match           0.3%;  Score 18;  DB 19;  Length 810;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4528 ttttagagcaaattgatc 4545
           ||||||||||||||||||
Db    335 ttttagagcaaattgatc 352


RESULT 167
V11433/c
ID   V11433 standard; cDNA; 811 BP.
XX
AC   V11433;
XX
DT   23-JUL-1998  (first entry)
XX
DE   Human secreted protein clone AJ143_1 cDNA 3'-end.
XX
KW   Secreted protein; prevention; treatment; gene therapy; ds.
```



```
XX
OS   Homo sapiens.
XX
PN   WO9801554-A2.
XX
PD   15-JAN-1998.
XX
PF   07-JUL-1997;   97WO-US11876.
XX
PR   09-JUL-1996;   96US-0677231.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Bowman M,  Evans C,  Jacobs K,  LaVallie ER,  McCoy JM;
PI   Merberg D,  Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1998-110230/10.
DR   P-PSDB; W58849.
XX
PT   Secreted proteins and polynucleotides encoding them - useful to
PT   prevent, treat and ameliorate medical conditions
XX
PS   Claim 25; Page 65-66; 93pp; English.
XX
CC   V11432 and V11433 encode fragments of a novel secreted protein derived
CC   from clone AJ143_1 which was isolated from a human adult testes cDNA
CC   library. The protein can be used to prevent, treat or ameliorate a
CC   medical condition, while the polynucleotides can be used for gene
CC   therapy.
XX
SQ   Sequence 811 BP; 210 A; 210 C; 166 G; 191 T; 34 other;


  Query Match           0.3%;  Score 18;  DB 19;  Length 811;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3961 ggtttccaaaaatcttaa 3978
           ||||||||||||||||||
Db    448 GGTTTCCAAAAATCTTAA 431


RESULT 168
X30394
ID   X30394 standard; DNA; 840 BP.
XX
AC   X30394;
XX
DT   14-MAY-1999  (first entry)
XX
DE   DNA encoding a human secreted protein.
```



```
XX
KW   Secreted protein; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; foetal deficiency; blood disorder;
KW   CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW   renal disease; diabetes; inflammation; allergy; ischemic shock;
KW   Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW   prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9907891-A1.
XX
PD   18-FEB-1999.
XX
PF   04-AUG-1998;   98WO-US16235.
XX
PR   19-AUG-1997;   97US-0056732.
PR   05-AUG-1997;   97US-0054798.
PR   05-AUG-1997;   97US-0054803.
PR   05-AUG-1997;   97US-0054804.
PR   05-AUG-1997;   97US-0054806.
PR   05-AUG-1997;   97US-0054807.
PR   05-AUG-1997;   97US-0054808.
PR   05-AUG-1997;   97US-0054809.
PR   05-AUG-1997;   97US-0055309.
PR   05-AUG-1997;   97US-0055310.
PR   05-AUG-1997;   97US-0055312.
PR   05-AUG-1997;   97US-0055386.
PR   05-AUG-1997;   97US-0055311.
PR   18-AUG-1997;   97US-0055970.
PR   18-AUG-1997;   97US-0055986.
PR   19-AUG-1997;   97US-0056365.
PR   19-AUG-1997;   97US-0056366.
PR   19-AUG-1997;   97US-0056557.
PR   19-AUG-1997;   97US-0056370.
PR   19-AUG-1997;   97US-0056371.
PR   19-AUG-1997;   97US-0056563.
PR   19-AUG-1997;   97US-0056731.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Ebner R,  Ferrie AM,  Greene JM,  Janat F,  Ni J;
PI   Olsen HS,  Rosen CA,  Ruben SM,  Soppet DR,  Young PE,  Yu G;
XX
DR   WPI; 1999-167452/14.
DR   P-PSDB; Y10873.
XX
PT   New isolated human genes encoding secreted polypeptides - useful for
PT   diagnosis and treatment of pathalogical diseases
XX
PS   Claim 3; Page 264; 331pp; English.
```



```
XX
CC   The specification describes secreted proteins and their corresponding
CC   polynucleotides which are useful for preventing, treating or ameliorating
CC   medical conditions, e.g. by protein or gene therapy. Pathological
CC   conditions can also be diagnosed by determining the amount of the
CC   secreted polypeptides in a sample or by determining the presence of
CC   mutations in the polynucleotides. Specific uses are described for each
CC   of the products, based on which tissues they are most highly
CC   expressed in, and include developing products for the diagnosis or
CC   treatment of cancer, tumours, neurodegenerative disorders, developmental
CC   abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC   diseases of the immune system, autoimmune diseases, hepatic and renal
CC   disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC   and cognitive disorders, schizophrenia, cardiovascular disorders,
CC   prostate diseases, asthma, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC   thymus, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
XX
SQ   Sequence 840 BP; 256 A; 129 C; 166 G; 287 T; 2 other;


  Query Match           0.3%;  Score 18;  DB 20;  Length 840;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    807 aatgaaaaaaaaaaaaaa 824


RESULT 169
X97814
ID   X97814 standard; cDNA; 855 BP.
XX
AC   X97814;
XX
DT   23-SEP-1999  (first entry)
XX
DE   Human secreted protein encoding cDNA #2.
XX
KW   Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
KW   diagnostic; gene therapy; chromosome mapping; secretion vector; ss.
XX
OS   Homo sapiens.
XX
PN   WO9925825-A2.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-IB01862.
```

```
XX
PR   04-SEP-1998;   98US-0099273.
PR   13-NOV-1997;   97US-0066677.
PR   17-DEC-1997;   97US-0069957.
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-347472/29.
DR   P-PSDB; Y36130.
XX
PT   Extended cDNAs encoding secreted proteins
XX
PS   Example 28; Page 148-149; 307pp; English.
XX
CC   X97813-X97906 represent extended cDNA's which encode novel human secreted
CC   proteins (see Y36129-Y36222) and which have cytostatic, thrombotic and
CC   osteopathic activity. The extended cDNAs can be used to express secreted
CC   proteins or parts of them or to obtain antibodies capable of binding to
CC   the secreted proteins. They may also be used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. Uses also include design
CC   of expression vectors and secretion vectors.
XX
SQ   Sequence 855 BP; 291 A; 137 C; 139 G; 288 T; 0 other;

  Query Match             0.3%;  Score 18;  DB 20;  Length 855;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     838 aatgaaaaaaaaaaaaaa 855


RESULT 170
Z07553
ID   Z07553 standard; cDNA; 861 BP.
XX
AC   Z07553;
XX
DT   03-DEC-1999  (first entry)
XX
DE   Maize glutathione-S-transferase encoding cDNA (clone CR1N.PK0014.B1).
XX
KW   Maize; Glutathione S-Transferase; GST; enzyme; transgenic plant;
KW   catalyse; detoxification; herbicide; xenobiotic; ss.
```



```
XX
OS   Zea mays.
XX
PN   US5962229-A.
XX
PD   05-OCT-1999.
XX
PF   05-SEP-1997;   97US-0924759.
XX
PR   05-SEP-1997;   97US-0924759.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   McGonigle B,  O'Keefe DP;
XX
DR   WPI; 1999-571265/48.
DR   P-PSDB; Y40567.
XX
PT   Maize Glutathione S-Transferase enzymes
XX
PS   Claim 2; Columns 43-44; 30pp; English.
XX
CC   The invention provides isolated nucleic acid sequences encoding maize
CC   Glutathione S-Transferase (GST) enzyme. The nucleic acid sequences are
CC   useful for producing transgenic plants which express increased levels of
CC   GST, the enzyme catalyzes the detoxification of herbicides and other
CC   xenobiotics. Sequences (Z07546-571) represent nucleotide sequences of
CC   various cDNA clones encoding maize GST enzymes (Y40560-571).
XX
SQ   Sequence 861 BP; 193 A; 291 C; 243 G; 134 T; 0 other;

  Query Match             0.3%;  Score 18;  DB 20;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     829 aatgaaaaaaaaaaaaaa 846


RESULT 171
Z94793
ID   Z94793 standard; cDNA; 861 BP.
XX
AC   Z94793;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Maize class III glutathione-S-transferase cDNA.
XX
```



```
KW   Maize; glutathione-S-transferase; GST; transgenic plant;
KW   xenobiotic; detoxification; herbicide tolerance; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             36..608
FT                   /*tag= a
FT                   /EC_number= 2.5.1.18
XX
PN   WO200018937-A1.
XX
PD   06-APR-2000.
XX
PF   30-SEP-1998;   98WO-US20502.
XX
PR   30-SEP-1998;   98WO-US20502.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   McGonigle B,  O'Keefe DP;
XX
DR   WPI; 2000-293169/25.
DR   P-PSDB; Y79397.
XX
PT   Novel maize glutathione-S-transferase enzymes used to alter the level
PT   of the enzyme in transformed plants and thus the level of
PT   detoxification of xenobiotic compounds
XX
PS   Claim 2; Page 49-50; 67pp; English.
XX
CC   The present sequence is that of the cDNA insert in maize clone
CC   crln.pk0014.g8, encoding a class III glutathione-S-transferase
CC   (GST, see Y79396).  The clone was isolated from a cDNA library
CC   prepared from light-grown 7-day seedling root tissue.  The
CC   invention provides maize GST enzymes involved in the detoxification
CC   of xenobiotic compounds in plants and seeds.  Chimeric genes
CC   encoding all or a portion of maize GST enzymes, host cells, and
CC   methods of recombinant production of maize GST enzymes are
CC   provided.  Methods of constructing transgenic plants having
CC   altered levels of GST enzymes, and screens for identifying maize
CC   GST enzyme substrates and maize GST inhibitors (useful as
CC   herbicides or herbicide synergists) are also provided.  The
CC   creation of transgenic plants in which GST is present at higher
CC   or lower levels than normal, or in cell types or developmental
CC   stages where it is not normally found, provides a means of
CC   modifying the herbicide tolerance of the plant.
XX
SQ   Sequence 861 BP; 193 A; 291 C; 243 G; 134 T; 0 other;
```



```
  Query Match             0.3%;  Score 18;  DB 21;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     829 aatgaaaaaaaaaaaaaa 846


RESULT 172
T72173
ID   T72173 standard; cDNA to mRNA; 882 BP.
XX
AC   T72173;
XX
DT   25-FEB-1998  (first entry)
XX
DE   Alzheimer's disease DNA sequence from plasmid pGCS1180.
XX
KW   Human; brain; Alzheimer's disease; diagnosis; antibody; expression; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             97..198
FT                   /*tag= a
FT                   /note= "Protein encoded by DNA from the human brain which
FT                         has different expression from the normal protein
FT                         when in an Alzheimer's disease patient"
XX
PN   WO9721807-A1.
XX
PD   19-JUN-1997.
XX
PF   12-DEC-1996;   96WO-JP03630.
XX
PR   12-DEC-1995;   95JP-0322745.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Hashida H,  Kuga T,  Nakagawa S,  Sakaki Y,  Zhao N;
XX
DR   WPI; 1997-332779/30.
DR   P-PSDB; W19950.
XX
PT   DNA sequences with varied expression in normal and Alzheimer
PT   patients - used for diagnosis and treatment of Alzheimer's disease
XX
PS   Claim 3; Page 48-49; 90pp; Japanese.
XX
CC   The present sequence represents a DNA fragment which is complementary to
```

```
CC   mRNA sequences whose level of expression in the brains of Alzheimer's
CC   disease patients varies from normal individuals. The present sequence
CC   represents a specifically claimed DNA fragment in which the expression
CC   is augmented to at least twice that in normal cases. The sequence, and
CC   antibodies to the proteins encoded by the sequence, may be used in the
CC   diagnosis, treatment and investigation of Alzheimer's disease.
XX
SQ   Sequence 882 BP; 306 A; 134 C; 142 G; 300 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 18;  Length 882;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    826 aatgaaaaaaaaaaaaaa 843


RESULT 173
Q14451
ID   Q14451 standard; DNA; 899 BP.
XX
AC   Q14451;
XX
DT   21-JAN-1992  (first entry)
XX
DE   Rat prolactin gene.
XX
KW   Recombinant DNA; ss.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             52..732
FT                   /*tag= a
XX
PN   JP03219876-A.
XX
PD   27-SEP-1991.
XX
PF   24-JAN-1990;   90JP-0014511.
XX
PR   24-JAN-1990;   90JP-0014511.
XX
PA   (SHIK-) SHIKISHIMA BOSEKI K.
XX
DR   WPI; 1991-329111/45.
DR   P-PSDB; R14599.
XX
PT   Mass-prodn. of rat prolactin-producing recombinant DNA - is by
```



```
PT   integration of promoter, Shine-Dalgarno sequence and translation
PT   initiation codon upstream of table 1 coding gene.
XX
PS   Claim 1; Table 1; 12pp; Japanese.
XX
CC   The gene can be ligated into an expression plasmid with a promoter,
CC   SD sequence and initiation codon for the prodn. of recombinant rat
CC   prolactin.  The protein is useful for the study of the physiological
CC   activity of rat and human prolactin.
CC   See also Q14452.
XX
SQ   Sequence 899 BP; 277 A; 206 C; 175 G; 241 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 12;  Length 899;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    867 aatgaaaaaaaaaaaaaa 884


RESULT 174
Q31989
ID   Q31989 standard; cDNA; 924 BP.
XX
AC   Q31989;
XX
DT   30-APR-1993  (first entry)
XX
DE   Guinea pig MBP-2 gene.
XX
KW   Major basic protein; anti-asthma; ss.
XX
OS   Cavia porcellus.
XX
FH   Key             Location/Qualifiers
FT   CDS             89..790
FT                   /*tag= a
FT   sig_peptide     89..134
FT                   /*tag= b
FT   sig_peptide     135..433
FT                   /*tag= c
FT                   /note= "precursor peptide"
FT   mat_peptide     434..790
FT                   /*tag= d
XX
PN   JP04320686-A.
XX
PD   11-NOV-1992.
```



```
XX
PF   22-APR-1991;   91JP-0090704.
XX
PR   22-APR-1991;   91JP-0090704.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
DR   WPI; 1992-426680/52.
DR   P-PSDB; R29717.
XX
PT   A guinea pig MBP gene - contains DNA base sequence coding MBP
PT   gene having 1 of 2 specified aminoacid sequences
XX
PS   Claim 1; Page 16; 20pp; Japanese.
XX
CC   The sequence shows the cDNA sequence of the guinea pig major basic
CC   protein (MBP)-2 gene. The sequence was obtd. by first isolating MBP-1
CC   and MBP-2 from the abdominal transudate of male guinea pigs treated
CC   with polymyxin and ascaris extract.  MBP-1 and MBP-2 were used to
CC   design probes for screening of a guinea pig eosinophil cDNA library.
CC   The MBPs may be prepd. in large amounts by this method and may be
CC   used for the confirmation of the effect of an anti-asthma drug.
CC   See also Q31988-93.
XX
SQ   Sequence 924 BP; 216 A; 236 C; 287 G; 185 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 13;  Length 924;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    896 aatgaaaaaaaaaaaaaa 913


RESULT 175
Q85268
ID   Q85268 standard; DNA; 940 BP.
XX
AC   Q85268;
XX
DT   16-AUG-1995  (first entry)
XX
DE   Spider venom FIL-705.
XX
KW   Spider venom; FIL-376; FIL-705; insecticide; pesticide;
KW   biological control agent; baculovirus; Lepidoptera; ds.
XX
OS   Filistrata hibernalis.
XX
```



```
FH   Key             Location/Qualifiers
FT   misc_difference 1..14
FT                   /*tag= a
FT                   /note= "bases at positions 1-14 are not given in
FT                         the specification"
FT   CDS             1..750
FT                   /*tag= b
XX
PN   WO9501996-A.
XX
PD   19-JAN-1995.
XX
PF   07-JUL-1994;   94WO-US07595.
XX
PR   07-JUL-1993;   93US-0089998.
XX
PA   (FMCC ) FMC CORP.
PA   (NPSP-) NPS PHARM INC.
XX
PI   Jackson JRH,  Kral RM,  Krapcho KJ;
XX
DR   WPI; 1995-066867/09.
DR   P-PSDB; R70761.
XX
PT   New insecticidal proteins from spider venom - and related
PT   expression vectors and transformed hosts, partic. for control of
PT   Lepidoptera larvae.
XX
PS   Disclosure; Page 42-43; 64pp; English.
XX
CC   Oligonucleotides corresp. to the partial sequence of insecticidal
CC   protein FIL-376 were used as primers for PCR amplification of
CC   Filistrata hibernalis cephalothorax cDNA.  Products were
CC   inserted into pKS and used to transform E. coli.  Subclone
CC   FIL-705 encoded a 249-amino acid protein (given in R70761) that
CC   may be a precursor of FIL-376. Expression of such proteins e.g.
CC   from baculovirus vectors is used for biological control of
CC   Lepidopteran pests.
XX
SQ   Sequence 940 BP; 322 A; 152 C; 193 G; 259 T; 14 other;


  Query Match             0.3%;  Score 18;  DB 16;  Length 940;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    908 aatgaaaaaaaaaaaaaa 925
```

```
RESULT 176
Z97102
ID   Z97102 standard; cDNA; 990 BP.
XX
AC   Z97102;
XX
DT   19-APR-2000  (first entry)
XX
DE   Human secreted protein gene 84 cDNA clone HERAD40, SEQ ID NO:94.
XX
KW   Human; secreted protein; cancer; tumour; developmental abnormality;
KW   foetal deficiency; blood disorder; immune system disorder; inflammation;
KW   autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW   schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW   atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW   digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
KW   therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO9966041-A1.
XX
PD   23-DEC-1999.
XX
PP   15-JUN-1999;   99WO-US13418.
XX
PR   16-JUN-1998;   98US-0089507.
PR   16-JUN-1998;   98US-0089508.
PR   16-JUN-1998;   98US-0089509.
PR   16-JUN-1998;   98US-0089510.
PR   22-JUN-1998;   98US-0090112.
PR   22-JUN-1998;   98US-0090113.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Rosen CA,  Wei Y,  Young PE,  Florence KA;
PI   Soppet DR,  Brewer LA,  Endress GA,  Carter KC,  Mucenski M,  Ebner R;
PI   Lafleur DW,  Olsen HS,  Shi Y,  Moore PA,  Komatsoulis G;
XX
DR   WPI; 2000-106100/09.
DR   P-PSDB; Y86298.
XX
PT   New isolated human genes and the secreted polypeptides they encode,
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 367; 586pp; English.
XX
CC   Z97019 to Z97137 represent 94 isolated human secreted protein genes.
CC   Y86215 to Y86333 are the secreted proteins encoded by the 94 human genes.
CC   This sequence represents a fragment of one of the human secreted
```



```
CC   proteins. The genes and their corresponding secreted polypeptides are
CC   useful for preventing, treating or ameliorating medical conditions,
CC   e.g., by protein or gene therapy. Also pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new genes. Specific
CC   uses are described for each of the 94 genes, based on which tissues they
CC   are most highly expressed in, and include developing products for the
CC   diagnosis or treatment of cancer, tumours, developmental abnormalities
CC   and foetal deficiencies, blood disorders, diseases of the immune system,
CC   autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC   disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC   disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC   disorders, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
CC   The sequences shown in Y86334 to Y86585 represent fragments of the
CC   secreted proteins.
XX
SQ   Sequence 990 BP; 273 A; 259 C; 191 G; 261 T; 6 other;


  Query Match             0.3%;  Score 18;  DB 21;  Length 990;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgct 406
          ||||||||||||||||||
Db    806 aaaaaaaaaaaaaatgct 823


RESULT 177
X58797
ID   X58797 standard; cDNA; 1002 BP.
XX
AC   X58797;
XX
DT   16-AUG-1999  (first entry)
XX
DE   Human leukaemia-associated gene 1 (Leul) cDNA.
XX
KW   Leul gene; leukaemia-associated gene 1; human; B-CLL;
KW   B-cell chronic lymphocytic leukaemia; tumour suppressor;
KW   diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             267..485
FT                   /*tag= a
FT   exon            1..385
FT                   /*tag= b
FT                   /number= 1
```



```
FT   exon            386..1002
FT                   /*tag= c
FT                   /number= 2
FT   polyA_signal    962..967
FT                   /*tag= d
XX
PN   WO9925736-A1.
XX
PD   27-MAY-1999.
XX
PP   13-NOV-1998;   98WO-SE02052.
XX
PR   13-NOV-1997;   97SE-0004162.
XX
PA   (KARO-) KAROLINSKA INNOVATIONS AB.
XX
PI   Einhorn S,  Grander D,  Liu Y,  Oscier D,  Rasool O;
PI   Zabarovsky E;
XX
DR   WPI; 1999-337986/28.
DR   P-PSDB; Y06218.
XX
PT   Chronic lymphocytic leukemia nucleotides and proteins
XX
PS   Claim 5; Page 58-59; 81pp; English.
XX
CC   This sequence represents human Leul (leukaemia associated gene 1)
CC   cDNA.  Previous studies had indicated the presence of a putative
CC   tumour suppressor gene on chromosome 13q14, commonly deleted in
CC   patients with B-cell chronic lymphocytic leukaemia (B-CLL).  In
CC   the present invention, the minimally deleted region has been
CC   narrowed to a 10 kb area, and 2 neighboring genes, termed Leul
CC   and Leu2 (see X58801), that span this region have been identified,
CC   with several patients showing deletion borders within these
CC   genes.  The Leul and Leu2 genes show little homology to previously
CC   published genes at the nucleotide and expected translated amino
CC   acid sequence level.  Mutational analysis of the Leul and Leu2
CC   genes in 170 CLL samples revealed no small intragenic mutations or
CC   point mutations.  However, in all cases of 13q14 loss examined,
CC   the first exons of both genes, which are only 300 bp apart, were
CC   deleted.  Thus, Leul and Leu2 are strong candidate tumour
CC   suppressor genes involved in B-CLL leukaemogenesis.  The invention
CC   relates to B-CLL associated nucleic acid and amino acid sequences,
CC   and to the use of the nucleic acids in gene therapy and for the
CC   production of proteins, which are used for the development of drugs
CC   against leukaemia and other malignancies of the same genetic area,
CC   for raising antibodies, and in diagnostic kits.
XX
SQ   Sequence 1002 BP; 288 A; 207 C; 216 G; 291 T; 0 other;
```



```
  Query Match             0.3%;  Score 18;  DB 20;  Length 1002;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db    977 aatgaaaaaaaaaaaaaa 994


RESULT 178
A08124
ID   A08124 standard; DNA; 1047 BP.
XX
AC   A08124;
XX
DT   27-JUN-2000  (first entry)
XX
DE   C. trachomatis MOMP containing fusion protein nucleotide sequence #5.
XX
KW   Chlamydia trachomatis; fusion protein; major outer membrane protein;
KW   MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;
KW   infection; infectious disease; ds.
XX
OS   Chlamydia trachomatis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1047
FT                   /*tag= a
FT                   /note= "Fusion protein containing at least part of a
FT                         major outer membrane protein (MOMP) of Chlamydia
FT                         trachomatis; no stop codon given"
XX
PN   JP2000041678-A.
XX
PD   15-FEB-2000.
XX
PF   28-JUL-1998;   98JP-0213212.
XX
PR   28-JUL-1998;   98JP-0213212.
XX
PA   (ELED ) DENKI KAGAKU KOGYO KK.
XX
DR   WPI; 2000-295780/26.
DR   P-PSDB; Y82392.
XX
PT   A soluble fused protein useful for diagnosis of Chlamydia infection,
PT   comprises at least part of major outer membrane protein (MOMP) of
PT   Chlamydia trachomatis
XX
PS   Claim 20; Page 29-30; 37pp; Japanese.
```

```
XX
CC   The present invention describes fusion proteins (I) comprising at least
CC   part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,
CC   at least one hydrophilic polypeptide having no immunoreactivity to
CC   human serum and their connected part. A08120 to A08125 encode
CC   specifically claimed examples of the fusion proteins given in Y82388 to
CC   Y82393. Also described is a method (A) for the detection of Chlamydia
CC   trachomatis antibody using (I) as the antigen. (I) is used for the
CC   diagnosis of Chlamydia trachomatis infectious diseases. The method can
CC   diagnose Chlamydia trachomatis infectious diseases specifically in a
CC   high sensitivity.
XX
SQ   Sequence 1047 BP; 315 A; 209 C; 233 G; 290 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 21;  Length 1047;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3171 caataacaataacaataa 3188
         ||||||||||||||||||
Db    504 caataacaataacaataa 521


RESULT 179
Q04287
ID   Q04287 standard; DNA; 1051 BP.
XX
AC   Q04287;
XX
DT   17-SEP-1990  (first entry)
XX
DE   Sequence encoding bovine trophoblast protein-1 (bTP-1) derived from
DE   clone lambda bTP509.
XX
KW   Bovine trophoblast protein-1; bTP-1; fertility; ds.
XX
OS   Sus scrofa.
XX
PN   EP367063-A.
XX
PD   09-MAY-1990.
XX
PF   23-OCT-1989;   89EP-0119642.
XX
PR   26-OCT-1988;   88US-0262870.
XX
PA   (UMOR ) UNIV OF MISSOURI.
XX
PI   Roberts MR,  Imakawa K;
XX
```



```
DR   WPI; 1990-141062/19.
XX
PT   Recombinant bovine trophoblast protein-1 -
PT   used for enhancing fertility or treating viral diseases in
PT   mammals, esp. cattle.
XX
PS   Disclosure; ; p; English.
XX
CC   The bTP-1 produced from the gene may be used to promote fertility
CC   or treat viral disease in cattle. The gene may also be used to
CC   provide transgenic animals with enhanced fertility, or in
CC   prophylactic and therapuetic treatment of other mammals.
CC   See also R04286 and R04288.
XX
SQ   Sequence 1051 BP; 291 A; 258 C; 228 G; 274 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 11;  Length 1051;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1030 aatgaaaaaaaaaaaaaa 1047


RESULT 180
Q04285
ID   Q04285 standard; cDNA; 1057 BP.
XX
AC   Q04285;
XX
DT   17-SEP-1990  (first entry)
XX
DE   cDNA clone of sequence encoding bovine trophoblast protein-1 (bTP-1).
XX
KW   Bovine trophoblast protein-1; bTP-1; fertility; ds.
XX
OS   Sus scrofa.
XX
FH   Key             Location/Qualifiers
FT   CDS             84..667
FT                   /*tag= a
XX
PN   EP367063-A.
XX
PD   09-MAY-1990.
XX
PF   23-OCT-1989;   89EP-0119642.
XX
PR   26-OCT-1988;   88US-0262870.
```



```
XX
PA   (UMOR ) UNIV OF MISSOURI.
XX
PI   Roberts MR,  Imakawa K;
XX
DR   WPI; 1990-141062/19.
DR   P-PSDB; R04539.
XX
PT   Recombinant bovine trophoblast protein-1 -
PT   used for enhancing fertility or treating viral diseases in
PT   mammals, esp. cattle.
XX
PS   Claim 3; Fig 1; 27pp; English.
XX
CC   The bTP-1 produced from the gene may be used to promote fertility
CC   or treat viral disease in cattle. The gene may also be used to
CC   provide transgenic animals with enhanced fertility, or in
CC   prophylactic and therapuetic treatment of other mammals.
XX
SQ   Sequence 1057 BP; 294 A; 260 C; 229 G; 274 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 11;  Length 1057;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1033 aatgaaaaaaaaaaaaaa 1050


RESULT 181
Z00410
ID   Z00410 standard; cDNA; 1080 BP.
XX
AC   Z00410;
XX
DT   04-OCT-1999  (first entry)
XX
DE   Human secreted protein cDNA endoding gene 1.
XX
KW   Secreted protein; human; treatment; diagnosis; therapy; cancer; tumour;
KW   neurodegenerative disorder; developmental abnormality; blood disorder;
KW   fetal deficiency; blood disorder; leukemia; immune system; inflammation;
KW   autoimmune disease; hepatic disease; renal disease; allergy; restenosis;
KW   ischaemic shock; Alzheimer's disease; cognitive disorder; schizophrenia;
KW   cardiovascular disorder; wound healing; stroke; arthritis; obesity;
KW   asthma; sepsis; acne; psoriasis; transplant rejection; infection; AIDS;
KW   metabolic disorder; ss.
XX
OS   Homo sapiens.
```



```
XX
PN   WO9938881-A1.
XX
PD   05-AUG-1999.
XX
PF   27-JAN-1999;   99WO-US01621.
XX
PR   30-JAN-1998;   98US-0073170.
PR   30-JAN-1998;   98US-0073159.
PR   30-JAN-1998;   98US-0073160.
PR   30-JAN-1998;   98US-0073161.
PR   30-JAN-1998;   98US-0073162.
PR   30-JAN-1998;   98US-0073164.
PR   30-JAN-1998;   98US-0073165.
PR   30-JAN-1998;   98US-0073167.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Carter KC,  Endress GA,  Feng P,  Ferrie AM,  Florence C;
PI   Florence KA,  Janat F,  Ni J,  Rosen CA,  Ruben SM;
PI   Soppet DR,  Young P,  Yu G;
XX
DR   WPI; 1999-469315/39.
DR   P-PSDB; Y25711, Y25779.
XX
PT   New isolated human genes and the secreted polypeptides they encode
PT   useful in, e.g. treatment of Alzheimer's
XX
PS   Claim 1a; Page 255-256; 393pp; English.
XX
CC   This invention describes novel human genes (see Z00410-Z00477) and the
CC   secreted proteins (see Y25711-Y25778) and fragments (see Y25779-Y25907)
CC   they encode. The polynucleotides and their corresponding secreted
CC   polypeptides are useful for preventing, treating or ameliorating medical
CC   conditions e.g. by protein or gene therapy. Also pathological conditions
CC   can be diagnosed by determining the amount of the new polypeptides in a
CC   sample or by determining the presence of mutations in the new
CC   polynucleotides. Specific uses are described for each of the 67
CC   polynucleotides of the invention, based on which tissues they are most
CC   highly expressed in, and include developing products for the diagnosis or
CC   treatment of cancer, tumours, neurodegenerative disorders, developmental
CC   abnormalities and fetal deficiencies, blood disorders, leukemias,
CC   diseases of the immune system, autoimmune diseases, hepatic and renal
CC   disease, inflammation, allergies, ischaemic shock, Alzheimer's and
CC   cognitive disorders, schizophrenia, restenosis, cardiovascular disorders,
CC   wound healing, stroke, arthritis, obesity, asthma, sepsis, acne,
CC   psoriasis, transplant rejection, metabolic disorders, infections and
CC   AIDS. The polypeptides are also useful for identifying their binding
CC   partners.
XX
SQ   Sequence 1080 BP; 289 A; 234 C; 303 G; 254 T; 0 other;
```

```
  Query Match           0.3%;  Score 18;  DB 20;  Length 1080;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1034 aatgaaaaaaaaaaaaaa 1051


RESULT 182
X08684/c
ID   X08684 standard; cDNA; 1083 BP.
XX
AC   X08684;
XX
DT   27-SEP-1999  (first entry)
XX
DE   Novel nucleotide sequence encoding new protein (Clone AJ143_1).
XX
KW   Polynucleotide; protein; nutrition; cytokine; cell proliferation;
KW   cell differentiation; immunostimulation; immunosuppression;
KW   haematopoiesis regulation; tissue growth; activin; inhibin;
KW   chemotaxis; chemokinesis; haemostasis; thrombolysis; receptor;
KW   ligand; anti-inflammatory; tumour suppression; gene therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             827..997
FT                   /*tag=  a
FT                   /product= "Novel protein"
XX
PN   WO9920644-A1.
XX
PD   29-APR-1999.
XX
PF   16-OCT-1998;   98WO-US22034.
XX
PR   10-OCT-1997;   97US-0955557.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Bowman MR,  Evans C,  Jacobs K,  Lavallie ER;
PI   McCoy JM,  Merberg D,  Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-288272/24.
DR   P-PSDB; W85719.
XX
PT   New polynucleotides encoding secreted human proteins
```



```
XX
PS   Claim 17; Page 103-104; 136pp; English.
XX
CC   The new human secreted proteins are encoded by polynucleotides
CC   obtained from human placenta, adult testes, fetal kidney, fetal
CC   brain, adult brain, adult brain and adult blood cDNA libraries.
CC   The polynucleotides and proteins are predicted to have biological
CC   activities which would make them suitable for treating, preventing or
CC   ameliorating medical conditions in humans and animals. Suggested
CC   activities include nutritional activity, cytokine and cell
CC   proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, and tumour inhibition
CC   activity.  The polynucleotides are also stated to be useful for gene
CC   therapy. The sequences are identified by a secretory leader
CC   sequence motif in the polynucleotide and it is thought that the
CC   encoded proteins have biological activity by virtue of their secreted
CC   nature. This clone was designated AJ143_1. A probe for this clone is
CC   described in X08699.
XX
SQ   Sequence 1083 BP; 284 A; 300 C; 246 G; 215 T; 38 other;


  Query Match           0.3%;  Score 18;  DB 20;  Length 1083;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3961 ggtttccaaaaatcttaa 3978
          ||||||||||||||||||
Db    663 GGTTTCCAAAAATCTTAA 646


RESULT 183
X60109
ID   X60109 standard; DNA; 1100 BP.
XX
AC   X60109;
XX
DT   05-AUG-1999  (first entry)
XX
DE   DNA encoding a formate transport associated protein, FTAP1.
XX
KW   Formate transport associated protein; FTAP1; formate production;
KW   gram-positive microorganism; formate transport; formate utilization;
KW   formate cycling; FTAP2; PurU; FMD; large scale fermentation; ss.
XX
OS   Bacillus subtilis.
XX
```



```
PN   WO9927107-A2.
XX
PD   03-JUN-1999.
XX
PF   19-NOV-1998;   98WO-US24873.
XX
PR   20-NOV-1997;   97GB-0024627.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Diaz-Torres M,  Miller B;
XX
DR   WPI; 1999-357843/30.
DR   P-PSDB; Y16106.
XX
PT   Regulation of formate production in gram-positive microorganisms
XX
PS   Disclosure; Fig 1A-C; 58pp; English.
XX
CC   The present sequence encodes a formate transport associated protein 1
CC   (FTAP1). The specification describes a method for modulating formate
CC   production levels in gram-positive microorganisms. Formate production
CC   levels in gram-positive microorganisms can be regulated by the modulation
CC   of four Bacillus proteins which appear to be associated with formate
CC   transport, utilization and cycling. These proteins are FTAP1 FTAP2, PurU
CC   and FMD. This also allows for increased yields of recombinant proteins
CC   being produced by the microorganism. The methods of the invention can
CC   be used to increase recombinant protein production in gram-positive
CC   microorganisms. The protein produced by a recombinant gram-positive host
CC   cell comprising a FTAP1, FTAP2, PurU and FMD will be secreted into the
CC   cell culture, from which it can be purified. The host cell can also be
CC   used to produce FTAP1, FTAP2, PurU and FMD proteins, preferably under
CC   large scale fermentation conditions. FTAP1, FTAP2, PurU and FMD
CC   polynucleotides can be used to detect the presence of gram-positive
CC   microorganism polynucleotide homologues through hybridization techniques
CC   and PCR. The polynucleotides also serve as a source of primers and
CC   probes.
XX
SQ   Sequence 1100 BP; 280 A; 205 C; 269 G; 346 T; 0 other;


  Query Match           0.3%;  Score 18;  DB 20;  Length 1100;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    403 tgcttgatattgttgtat 420
          ||||||||||||||||||
Db     62 tgcttgatattgttgtat 79


RESULT 184
```



```
X00711
ID   X00711 standard; DNA; 1134 BP.
XX
AC   X00711;
XX
DT   25-MAR-1999  (first entry)
XX
DE   Human secreted protein gene 48 clone HTEJN13.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9842738-A1.
XX
PD   01-OCT-1998.
XX
PF   19-MAR-1998;   98WO-US05311.
XX
PR   30-MAY-1997;   97US-0050937.
PR   21-MAR-1997;   97US-0041276.
PR   21-MAR-1997;   97US-0041277.
PR   21-MAR-1997;   97US-0041281.
PR   21-MAR-1997;   97US-0042344.
PR   30-MAY-1997;   97US-0048069.
PR   30-MAY-1997;   97US-0048094.
PR   30-MAY-1997;   97US-0048095.
PR   30-MAY-1997;   97US-0048096.
PR   30-MAY-1997;   97US-0048099.
PR   30-MAY-1997;   97US-0048131.
PR   30-MAY-1997;   97US-0048135.
PR   30-MAY-1997;   97US-0048154.
PR   30-MAY-1997;   97US-0048160.
PR   30-MAY-1997;   97US-0048186.
PR   30-MAY-1997;   97US-0048187.
PR   30-MAY-1997;   97US-0048188.
PR   30-MAY-1997;   97US-0048350.
PR   30-MAY-1997;   97US-0048351.
PR   30-MAY-1997;   97US-0048352.
PR   30-MAY-1997;   97US-0048355.
PR   05-AUG-1997;   97US-0054804.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
```

PI Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
PI Rosen CA, Ruben SM, Shi Y, Young P;
XX
DR WPI; 1999-070066/06.
DR P-PSDB; W67907.
XX
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 260; 385pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. X00602) for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 87 novel genes and their fragments (nucleic acid
CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 87 polynucleotides, based on
CC which tissues they are most highly expressed in (see X00611 for described
CC uses).
XX
SQ Sequence 1134 BP; 325 A; 271 C; 256 G; 270 T; 12 other;

Query Match 0.3%; Score 18; DB 20; Length 1134;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1117 aatgaaaaaaaaaaaaaa 1134

RESULT 185
X00658
ID X00658 standard; DNA; 1147 BP.
XX
AC X00658;
XX
DT 25-MAR-1999 (first entry)
XX
DE Human secreted protein gene 48 clone HTEJN13.
XX

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9842738-A1.
XX
PD 01-OCT-1998.
XX
PF 19-MAR-1998; 98WO-US05311.
XX
PR 30-MAY-1997; 97US-0050937.
PR 21-MAR-1997; 97US-0041276.
PR 21-MAR-1997; 97US-0041277.
PR 21-MAR-1997; 97US-0041281.
PR 21-MAR-1997; 97US-0042344.
PR 30-MAY-1997; 97US-0048069.
PR 30-MAY-1997; 97US-0048094.
PR 30-MAY-1997; 97US-0048095.
PR 30-MAY-1997; 97US-0048096.
PR 30-MAY-1997; 97US-0048099.
PR 30-MAY-1997; 97US-0048131.
PR 30-MAY-1997; 97US-0048135.
PR 30-MAY-1997; 97US-0048154.
PR 30-MAY-1997; 97US-0048160.
PR 30-MAY-1997; 97US-0048186.
PR 30-MAY-1997; 97US-0048187.
PR 30-MAY-1997; 97US-0048188.
PR 30-MAY-1997; 97US-0048350.
PR 30-MAY-1997; 97US-0048351.
PR 30-MAY-1997; 97US-0048352.
PR 30-MAY-1997; 97US-0048355.
PR 05-AUG-1997; 97US-0054804.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
PI Rosen CA, Ruben SM, Shi Y, Young P;
XX
DR WPI; 1999-070066/06.
DR P-PSDB; W67854.
XX
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 210-211; 385pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. X00602) for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 87 novel genes and their fragments (nucleic acid
CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 87 polynucleotides, based on
CC which tissues they are most highly expressed in (see X00611 for described
CC uses).
XX
SQ Sequence 1147 BP; 338 A; 275 C; 264 G; 270 T; 0 other;

Query Match 0.3%; Score 18; DB 20; Length 1147;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1120 aatgaaaaaaaaaaaaaa 1137

RESULT 186
X30152
ID X30152 standard; DNA; 1197 BP.
XX
AC X30152;
XX
DT 18-JUN-1999 (first entry)
XX
DE Human secreted protein gene 8.
XX
KW Human; secreted protein; cancer; tumour; developmental abnormality;
KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS; ss.
XX
OS Homo sapiens.
XX

PN WO9910363-A1.
XX
PD 04-MAR-1999.
XX
PF 27-AUG-1998; 98WO-US17709.
XX
PR 29-AUG-1997; 97US-0056271.
PR 29-AUG-1997; 97US-0056073.
PR 29-AUG-1997; 97US-0056247.
PR 29-AUG-1997; 97US-0056270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
XX
DR WPI; 1999-190585/16.
DR P-PSDB; Y04300.
XX
PT New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 136; 170pp; English.
XX
CC X30145 to X30173 represent 29 isolated human secreted protein genes.
CC Y04293 to Y04321 represent the secreted proteins encoded by the 29 human
CC genes. The genes and their corresponding secreted polypeptides are
CC useful for preventing, treating or ameliorating medical conditions,
CC e.g. by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new genes. Specific
CC uses are described for each of the 29 genes, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC disorders, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The sequences given in X30174 to X30182 and Y04322 to Y04334 are used
CC in the exemplification of the present invention.
XX
SQ Sequence 1197 BP; 289 A; 270 C; 287 G; 345 T; 6 other;

Query Match 0.3%; Score 18; DB 20; Length 1197;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402

```
                  ||||||||||||||||||
Db    1160 aatgaaaaaaaaaaaaaa 1177


RESULT 187
A37072
ID   A37072 standard; cDNA; 1204 BP.
XX
AC   A37072;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1279 (UNQ649) cDNA sequence SEQ ID NO:169.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20111.
XX
PR   01-SEP-1998;   98US-0098716.
PR   01-SEP-1998;   98US-0098749.
PR   01-SEP-1998;   98US-0098750.
PR   02-SEP-1998;   98US-0098803.
PR   02-SEP-1998;   98US-0098821.
PR   02-SEP-1998;   98US-0098843.
PR   09-SEP-1998;   98US-0099536.
PR   09-SEP-1998;   98US-0099596.
PR   09-SEP-1998;   98US-0099598.
PR   09-SEP-1998;   98US-0099602.
PR   09-SEP-1998;   98US-0099642.
PR   10-SEP-1998;   98US-0099741.
PR   10-SEP-1998;   98US-0099754.
PR   10-SEP-1998;   98US-0099763.
PR   10-SEP-1998;   98US-0099792.
PR   10-SEP-1998;   98US-0099808.
PR   10-SEP-1998;   98US-0099812.
PR   10-SEP-1998;   98US-0099815.
PR   10-SEP-1998;   98US-0099816.
PR   15-SEP-1998;   98US-0100385.
PR   15-SEP-1998;   98US-0100388.
PR   15-SEP-1998;   98US-0100390.
PR   16-SEP-1998;   98US-0100584.
PR   16-SEP-1998;   98US-0100627.
PR   16-SEP-1998;   98US-0100661.
```



```
PR   16-SEP-1998;   98US-0100662.
PR   16-SEP-1998;   98US-0100664.
PR   17-SEP-1998;   98US-0100683.
PR   17-SEP-1998;   98US-0100684.
PR   17-SEP-1998;   98US-0100710.
PR   17-SEP-1998;   98US-0100711.
PR   17-SEP-1998;   98US-0100919.
PR   17-SEP-1998;   98US-0100930.
PR   18-SEP-1998;   98US-0100848.
PR   18-SEP-1998;   98US-0100849.
PR   18-SEP-1998;   98US-0101014.
PR   18-SEP-1998;   98US-0101068.
PR   18-SEP-1998;   98US-0101071.
PR   22-SEP-1998;   98US-0101279.
PR   23-SEP-1998;   98US-0101471.
PR   23-SEP-1998;   98US-0101472.
PR   23-SEP-1998;   98US-0101474.
PR   23-SEP-1998;   98US-0101475.
PR   23-SEP-1998;   98US-0101476.
PR   23-SEP-1998;   98US-0101477.
PR   23-SEP-1998;   98US-0101479.
PR   24-SEP-1998;   98US-0101738.
PR   24-SEP-1998;   98US-0101741.
PR   24-SEP-1998;   98US-0101743.
PR   24-SEP-1998;   98US-0101915.
PR   24-SEP-1998;   98US-0101916.
PR   29-SEP-1998;   98US-0102207.
PR   29-SEP-1998;   98US-0102240.
PR   29-SEP-1998;   98US-0102307.
PR   29-SEP-1998;   98US-0102310.
PR   29-SEP-1998;   98US-0102331.
PR   30-SEP-1998;   98US-0102484.
PR   30-SEP-1998;   98US-0102487.
PR   30-SEP-1998;   98US-0102570.
PR   30-SEP-1998;   98US-0102571.
PR   01-OCT-1998;   98US-0102684.
PR   01-OCT-1998;   98US-0102687.
PR   02-OCT-1998;   98US-0102965.
PR   06-OCT-1998;   98US-0103258.
PR   06-OCT-1998;   98US-0103449.
PR   07-OCT-1998;   98US-0103314.
PR   07-OCT-1998;   98US-0103315.
PR   07-OCT-1998;   98US-0103328.
PR   07-OCT-1998;   98US-0103395.
PR   07-OCT-1998;   98US-0103396.
PR   07-OCT-1998;   98US-0103401.
PR   08-OCT-1998;   98US-0103633.
PR   08-OCT-1998;   98US-0103678.
PR   08-OCT-1998;   98US-0103679.
PR   08-OCT-1998;   98US-0103711.
PR   14-OCT-1998;   98US-0104257.
```



```
PR   20-OCT-1998;   98US-0104987.
PR   20-OCT-1998;   98US-0105000.
PR   20-OCT-1998;   98US-0105002.
PR   21-OCT-1998;   98US-0105104.
PR   22-OCT-1998;   98US-0105169.
PR   22-OCT-1998;   98US-0105266.
PR   26-OCT-1998;   98US-0105693.
PR   26-OCT-1998;   98US-0105694.
PR   27-OCT-1998;   98US-0105807.
PR   27-OCT-1998;   98US-0105881.
PR   27-OCT-1998;   98US-0105882.
PR   27-OCT-1998;   98US-0106062.
PR   28-OCT-1998;   98US-0106023.
PR   28-OCT-1998;   98US-0106029.
PR   28-OCT-1998;   98US-0106030.
PR   28-OCT-1998;   98US-0106032.
PR   28-OCT-1998;   98US-0106033.
PR   28-OCT-1998;   98US-0106178.
PR   29-OCT-1998;   98US-0106248.
PR   29-OCT-1998;   98US-0106384.
PR   29-OCT-1998;   98US-0108500.
PR   30-OCT-1998;   98US-0106464.
PR   03-NOV-1998;   98US-0106856.
PR   03-NOV-1998;   98US-0106902.
PR   03-NOV-1998;   98US-0106905.
PR   03-NOV-1998;   98US-0106919.
PR   03-NOV-1998;   98US-0106932.
PR   03-NOV-1998;   98US-0106934.
PR   10-NOV-1998;   98US-0107703.
PR   17-NOV-1998;   98US-0108775.
PR   17-NOV-1998;   98US-0108779.
PR   17-NOV-1998;   98US-0108787.
PR   17-NOV-1998;   98US-0108788.
PR   17-NOV-1998;   98US-0108801.
PR   17-NOV-1998;   98US-0108802.
PR   17-NOV-1998;   98US-0108806.
PR   17-NOV-1998;   98US-0108807.
PR   17-NOV-1998;   98US-0108867.
PR   17-NOV-1998;   98US-0108925.
PR   18-NOV-1998;   98US-0108848.
PR   18-NOV-1998;   98US-0108849.
PR   18-NOV-1998;   98US-0108850.
PR   18-NOV-1998;   98US-0108851.
PR   18-NOV-1998;   98US-0108852.
PR   18-NOV-1998;   98US-0108858.
PR   18-NOV-1998;   98US-0108904.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker K,  Goddard A,  Gurney AL,  Smith V,  Watanabe CK,  Wood WI;
XX
```



```
DR   WPI; 2000-237871/20.
DR   P-PSDB; Y99390.
XX
PT   New mammalian DNA sequences encoding transmembrane, receptor or
PT   secreted PRO polypeptides, useful for screening of potential peptide or
PT   small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS   Claim 2; Fig 101; 773pp; English.
XX
CC   A37022 to A37144 encode the new isolated human transmembrane, receptor
CC   or secreted PRO polypeptides given in Y99340 to Y99462. The
CC   transmembrane and receptor PRO proteins can be used for screening of
CC   potential peptide or small molecule inhibitors of the relevant
CC   receptor/ligand interactions. The polypeptides and nucleotide sequences
CC   encoding then have various industrial applications, including uses as
CC   pharmaceutical and diagnostic agents. A37145 to A37330 represent
CC   PCR primers and hybridisation probes used in the isolation of the PRO
CC   polypeptides from the present invention.
XX
SQ   Sequence 1204 BP; 306 A; 364 C; 294 G; 240 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 21;  Length 1204;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1163 aatgaaaaaaaaaaaaaa 1180


RESULT 188
Z33523
ID   Z33523 standard; cDNA; 1223 BP.
XX
AC   Z33523;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Human prostate cancer-associated EST 47.
XX
KW   Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
KW   cancer; tissue specificity; human; ss.
XX
OS   Homo sapiens.
XX
PN   DE19811194-A1.
XX
PD   16-SEP-1999.
XX
PF   10-MAR-1998;   98DE-1011194.
```

XX
PR 10-MAR-1998; 98DE-1011194.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX
DR WPI; 1999-519629/44.
DR P-PSDB; Y48412.
XX
PT New nucleic acid expressed at high level in normal prostatic tissue and
PT encoded polypeptides, used to treat cancer and screen for therapeutic
PT agents -
XX
PS Claim 3; 111; 194pp; German.
XX
CC This invention describes novel nucleic acid sequences (A) that are
CC expressed at high level in normal prostatic tissue. Polypeptides (I)
CC encoded by (A) are used: (a) for identifying agents for treatment of
CC prostatic cancer and (b) for therapy of prostate cancer, optionally
CC where expressed by gene therapy methods. (A) is also used to isolate
CC full-length genes (for gene therapy) and for recombinant production of
CC (I), which can be used to raise specific antibodies. (A) are identified
CC by assembly of ESTs (expressed sequence tags) before these are analyzed
CC for expression pattern (tissue specificity). This approach eliminates
CC many of the false results, as regards tissue specificity, associated with
CC known methods that use single (usually short) ESTs. Z33477-Z33540
CC represent expressed sequence tags described in the method of the
CC invention.
XX
SQ Sequence 1223 BP; 322 A; 358 C; 311 G; 232 T; 0 other;

Query Match 0.3%; Score 18; DB 20; Length 1223;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 386 atgaaaaaaaaaaaaaat 403
||||||||||||||||||
Db 1198 atgaaaaaaaaaaaaaat 1215

RESULT 189
T67029/c
ID T67029 standard; DNA; 1230 BP.
XX
AC T67029;
XX
DT 04-AUG-1997 (first entry)
XX
DE Salmonella secreted protein SspC DNA.

XX
KW Salmonella secreted protein; SspC; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
XX
OS Salmonella typhimurium.
XX
FH Key Location/Qualifiers
FT CDS 1..1230
FT /*tag= a
FT /product= SspC
FT /note= "the SspC amino acid sequence deduced
FT from the coding sequence does not fully
FT correspond to the SspC sequence given in
FT Fig 13"
XX
PN WO9718225-A1.
XX
PD 22-MAY-1997.
XX
PF 14-NOV-1996; 96WO-US18504.
XX
PR 14-NOV-1995; 95US-0006733.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Miller SI;
XX
DR WPI; 1997-289217/26.
DR P-PSDB; W15269.
XX
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
XX
PS Claim 5; Fig 20; 95pp; English.
XX
CC A DNA molecule (T67029) codes for Salmonella secreted protein SspC
CC (W15269), a virulence factor involved in bacterial mediated
CC endocytosis, which is able to translocate into human epithelial
CC cells. To identify Ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH were identified. Of these, one mutant lacked Ssp of 87,
CC 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region (see
CC also T67035) identified 2 complete and 2 partial open reading
CC frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC (T67031). These nucleic acids can be used in methods for diagnosing
CC Salmonella infection and in methods of targetting an antigen to
CC epithelial cells and of inducing a cytotoxic T cell immune response
CC in a mammal. Ssp gene mutation can be used to create attenuated

CC Salmonella strains useful as live vaccines.
XX
SQ Sequence 1230 BP; 362 A; 270 C; 317 G; 281 T; 0 other;

Query Match 0.3%; Score 18; DB 18; Length 1230;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4855 agaatatctttagcgcta 4872
||||||||||||||||||
Db 107 AGAATATCTTTAGCGCTA 90

RESULT 190
Z41296
ID Z41296 standard; cDNA; 1240 BP.
XX
AC Z41296;
XX
DT 18-JAN-2000 (first entry)
XX
DE Human normal ovarian tissue derived cDNA 75.
XX
KW Human; ovary; screening; ovarian cancer; treatment; ss.
XX
OS Homo sapiens.
XX
PN DE19816395-A1.
XX
PD 07-OCT-1999.
XX
PF 03-APR-1998; 98DE-1016395.
XX
PR 03-APR-1998; 98DE-1016395.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-552352/47.
XX
PT Nucleic acid sequences potentially useful in diagnosis or therapy of
PT ovarian cancer -
XX
PS Claim 3; Page 185; 274ppp; German.
XX
CC This invention describes novel nucleic acid sequences that are highly
CC expressed in normal ovary tissue. Artificial chromosomes and cosmid
CC clones containing the sequences can be used as gene transfer vehicles.
CC The sequences can be used to produce DNA fragments containing full-length

CC genes. Host cells transformed with the sequences can be used to produce
CC polypeptides or polypeptide fragments, which can be used to screen phage
CC displays for polypeptides that bind to them, or as tools for identifying
CC agents active against ovarian cancer, or to prepare medicaments for
CC treating ovarian cancer. The cDNA sequences can be used to obtain genomic
CC genes, their promoters, enhancers, silencers, exon structures, intron
CC structures and their splice variants. Z41222-Z41324 represent cDNA
CC sequences derived from normal human ovarian tissue and which encode the
CC protein fragments represented in Y59724-Y59837.
XX
SQ Sequence 1240 BP; 367 A; 240 C; 311 G; 322 T; 0 other;

Query Match 0.3%; Score 18; DB 20; Length 1240;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1214 aatgaaaaaaaaaaaaaa 1231

RESULT 191
V26294
ID V26294 standard; DNA; 1250 BP.
XX
AC V26294;
XX
DT 28-JUL-1998 (first entry)
XX
DE Phosphate starvation-induced protein kinase psrPK (psrl) encoding DNA.
XX
KW Phosphate starvation-induced protein kinase; psrPK; psrl; glucosidase;
KW phosphate transporter; phosphate deprivation; Arabidopsis thaliana;
KW Brassica nigra; photosynthetic organism; phosphorus metabolism; growth;
KW reproduction; metabolic content; flowering; drought; cold tolerance; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 90..1109
FT /*tag= a
FT /product= "phosphate starvation-induced protein kinase"
XX
PN WO9805760-A2.
XX
PD 12-FEB-1998.
XX
PF 30-JUL-1997; 97WO-CA00532.
XX
PR 31-JUL-1996; 96US-0688988.

```
PR   31-JUL-1996;   96CA-2182421.
XX
PA   (TOOH ) UNIV QUEENS KINGSTON.
XX
PI   Lefebvre DD,  Malboobi MA;
XX
DR   WPI; 1998-159169/14.
DR   P-PSDB; W56020.
XX
PT   New isolated phosphate starvation-inducible genes - used to develop
PT   plants or other photosynthetic organisms with altered phosphorus
PT   metabolism, e.g. to alter growth or reproduction.
XX
PS   Claim 2; Fig 5; 131pp; English.
XX
CC   The present sequence encodes a phosphate starvation-induced protein
CC   kinase from a photosynthetic organism, in which transcription of the
CC   DNA is induced by phosphate deficiency. The products and methods of
CC   the present invention can be used to modify the phosphorus metabolism
CC   in plants and other photosynthetic organisms, e.g. to alter growth,
CC   reproduction, metabolic content, flowering, drought or cold tolerance
CC   or nutritive value.
XX
SQ   Sequence 1250 BP; 428 A; 206 C; 269 G; 347 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 19;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1232 aatgaaaaaaaaaaaaaa 1249


RESULT 192
T43482/c
ID   T43482 standard; DNA; 1254 BP.
XX
AC   T43482;
XX
DT   11-AUG-1997  (first entry)
XX
DE   ATM gene exons 48-49.
XX
KW   ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
KW   central nervous system; immune system; chromosomal instability; therapy;
KW   cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW   multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW   general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW   blood vessel; bulbar conjunctiva; facial skin; A-T; ss.
```



```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   intron          1..103
FT                   /*tag= a
PT                   /number= 47
FT   exon            104..338
FT                   /*tag= b
FT                   /number= 48
FT   intron          339..858
FT                   /*tag= c
FT                   /number= 48
PT   exon            859..1026
FT                   /*tag= d
FT                   /number= 49
FT   intron          1027..1254
FT                   /*tag= e
FT                   /number= 49
XX
PN   WO9636691-A1.
XX
PD   21-NOV-1996.
XX
PF   16-MAY-1996;   96WO-US07025.
XX
PR   08-APR-1996;   96US-0629001.
PR   16-MAY-1995;   95US-0441822.
XX
PA   (KOHN/) KOHN K I.
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
PI   Shiloh Y;
XX
DR   WPI; 1997-012070/01.
XX
PT   New isolated ataxia-telangiectasia gene - used to develop prods. for
PT   the study, diagnosis and treatment of ataxia-telangiectasia.
XX
PS   Claim 1; Page 69; 153pp; English.
XX
CC   T43444-T43496 represent exons of the ATM gene of the invention.
CC   Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
CC   deletions in the coding region of the ATM gene. A-T is a progressive
CC   genetic disorder affecting the central nervous and immune systems. A-T
CC   involves chromosomal instability, cancer predisposition, radiation
CC   sensitivity, and cell cycle abnormalities. A-T is a multi-system disease
CC   inherited in an autosomal recessive manner. Cerebellar ataxia that
CC   gradually develops into general motor dysfunction is the first clinical
CC   hallmark of A-T, and results from progressive loss of Purkinje cells in
CC   the cerebellum. Oculocutaneous telangiectasia (dilation of blood vessels)
```



```
CC   develops in the bulbar conjunctiva and facial skin, and is later
CC   accompanied by graying of the hair and atrophic changes in the skin. The
CC   co-occurrence of cerebellar ataxia and telangiectases in the conjunctivae
CC   and occasionally on the facial skin (the second early hallmark of A-T)
CC   usually establishes the differential diagnosis of A-T from other
CC   cerebellar ataxias. The ATM gene, can be used in methods for detecting
CC   carriers of a defective gene that causes A-T. The gene can also be used
CC   to generate antibodies. The antibodies and methods can be used in the
CC   study, diagnosis and therapy of A-T.
XX
SQ   Sequence 1254 BP; 417 A; 199 C; 220 G; 418 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 18;  Length 1254;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      66 AATGAAAAAAAAAAAAAA 49


RESULT 193
Z33591
ID   Z33591 standard; cDNA; 1265 BP.
XX
AC   Z33591;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Human breast tumour-associated EST 51.
XX
KW   Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW   medicaments; gene therapy; treatment; fat metabolism; ss.
XX
OS   Homo sapiens.
XX
PN   DE19813835-A1.
XX
PD   23-SEP-1999.
XX
PF   20-MAR-1998;   98DE-1013835.
XX
PR   20-MAR-1998;   98DE-1013835.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E,  Rosenthal A;
XX
DR   WPI; 1999-528979/45.
XX
```



```
PT   Human nucleic acid sequences and protein products from normal breast
PT   tissue, useful for breast cancer therapy -
XX
PS   Claim 3; 139; 206pp; German.
XX
CC   This invention describes novel human nucleic acid sequences from normal
CC   breast tissue which have cytostatic activity. The nucleic acid sequences
CC   can be used to produce and isolate full-length gene sequences. They can
CC   be used to express proteins, which can be used as tools to find an
CC   activity against breast cancer. The sequences can be used in sense or
CC   antisense form. They are especially useful for medicaments for gene
CC   therapy to treat breast cancer and for treating illnesses associated
CC   with fat metabolism. Z33541-Z33610 represent expressed sequence tags
CC   described in the method of the invention.
XX
SQ   Sequence 1265 BP; 369 A; 252 C; 245 G; 399 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 20;  Length 1265;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1245 aatgaaaaaaaaaaaaaa 1262


RESULT 194
Q15023
ID   Q15023 standard; cDNA; 1290 BP.
XX
AC   Q15023;
XX
DT   16-MAR-1992  (first entry)
XX
DE   Rat prolactin coding sequence.
XX
KW   rPRL; plasmid prPRL1; ds.
XX
OS   Rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             441..1121
FT                   /*tag= a
FT   polyA_signal    1237..1242
FT                   /*tag= b
XX
PN   JP03247284-A.
XX
PD   05-NOV-1991.
XX
```

```
PF   26-FEB-1990;   90JP-0044804.
XX
PR   26-FEB-1990;   90JP-0044804.
XX
PA   (SHIK-) SHIKISHIMA BOSEKI K.
XX
DR   WPI; 1991-366331/50.
DR   P-PSDB; R15348.
XX
PT   Plasmid DNA for prodn. of rat prolactin - obtd. by bonding vector
PT   plasmid to terminal of DNA contg. specified base sequence
XX
PS   Claim 1; Table 1; 10pp; Japanese.
XX
CC   This is the first part of the sequence of prPRL1, up to the polyA
CC   tail. The precise number of adenyl residues present is not given,
CC   therefore 30 are shown here; the remainder of the prPRL1 sequence
CC   (Q15024) begins with 30 adenyl residues. The fragment of this
CC   sequence that includes the CDS, i.e. nucleotides 390 to 1259, is
CC   specified for use in the construction of a plasmid for the
CC   expression of rat prolactin.
XX
SQ   Sequence 1290 BP; 373 A; 331 C; 265 G; 321 T; 0 other;


  Query Match           0.3%;  Score 18;  DB 12;  Length 1290;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1256 aatgaaaaaaaaaaaaaa 1273


RESULT 195
Z42221
ID   Z42221 standard; cDNA; 1303 BP.
XX
AC   Z42221;
XX
DT   31-JAN-2000  (first entry)
XX
DE   Human normal bladder tissue cDNA derived EST 100.
XX
KW   Human; bladder; treatment; EST; expressed sequence tag; cytostatic;
KW   cancer; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   DE19818620-A1.
XX
```



```
PD   28-OCT-1999.
XX
PF   21-APR-1998;   98DE-1018620.
XX
PR   21-APR-1998;   98DE-1018620.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-602416/52.
DR   P-PSDB; Y60521, Y60522.
XX
PT   New polypeptides and their nucleic acids, useful for treatment of
PT   bladder tumour and identification of therapeutic agents
XX
PS   Claim 3; Page 221-222; 366pp; German.
XX
CC   This invention describes novel polypeptide fragment sequences (I) and
CC   their encoding nucleic acids (II) which are highly expressed in normal
CC   bladder tissue and have cytostatic activity. (II) are used for
CC   recombinant expression of (I) and to isolate complete genes. (I) are
CC   used to identify agents suitable for the treatment of bladder tumours, to
CC   directly treat this form of cancer (including expression from gene
CC   therapy vectors), or are used in a preparation for cancer treatment. (I)
CC   is also used for the generation of specific antibodies. (II) are
CC   identified by assembling ESTs (expressed sequence tags) from a particular
CC   tissue type before comparison of expression patterns. This allows a
CC   significantly longer fragment of the gene to be revealed, and therefore
CC   reduces the number of failures because of ESTs from different libraries
CC   representing different parts of the same unknown gene distorting the
CC   estimated frequency of occurrence in a particular tissue. Z42122-Z42248
CC   represent EST fragments derived from a human normal bladder tissue cDNA
CC   library which encode the protein fragments represented in Y60329-Y60591.
XX
SQ   Sequence 1303 BP; 438 A; 200 C; 227 G; 438 T; 0 other;


  Query Match           0.3%;  Score 18;  DB 20;  Length 1303;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1270 aatgaaaaaaaaaaaaaa 1287


RESULT 196
Z30222
ID   Z30222 standard; cDNA; 1314 BP.
XX
```



```
AC   Z30222;
XX
DT   11-FEB-2000  (first entry)
XX
DE   cDNA encoding a human prostate-associated serum protease (PRASP).
XX
KW   Human; prostate-associated serum protease; PRASP; neuropsin; PSA;
KW   Incyte clone 2723646; reproductive disorder; cancer;
KW   abnormal prolactin production; infertility; tubal disease;
KW   ovulatory defect; endometriosis; polycystic ovary syndrome;
KW   autoimmune disorder; ectopic pregnancy; breast cancer;
KW   abnormal spermatogenesis; testicular cancer; adenocarcinoma; leukaemia;
KW   lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..976
FT                   /*tag= a
FT                   /product= "prostate-associated serum protease"
FT   misc_feature    146..181
FT                   /*tag= b
FT                   /note= "these nucleotides are separately claimed
FT                        under claim 10"
FT   misc_feature    344..382
FT                   /*tag= c
FT                   /note= "these nucleotides are separately claimed
FT                        under claim 10"
FT   misc_feature    551..589
FT                   /*tag= d
FT                   /note= "these nucleotides are separately claimed
FT                        under claim 10"
XX
PN   WO9941387-A2.
XX
PD   19-AUG-1999.
XX
PF   05-FEB-1999;   99WO-US02571.
XX
PR   17-FEB-1998;   98US-0025059.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Yang YT,  Corley NC,  Guegler KJ;
XX
DR   WPI; 2000-012993/01.
DR   P-PSDB; Y43636.
XX
PT   New prostate-associated serum protease and polynucleotides which
PT   identify and encode PRASP, useful for treating reproductive disorders
PT   and cancer
```



```
XX
PS   Claim 7; Fig 1A-D; 67pp; English.
XX
CC   The present sequence encodes a human prostate-associated serum protease
CC   (PRASP). The protein shows homology to neuropsin, a brain-specific
CC   protease in mice, and PSA, a prostate-specific protease in humans.
CC   Nucleic acids encoding PRASP were first identified in Incyte clone
CC   2723646 from the lung tumour cDNA library. Pharmaceutical compositions
CC   containing PRASP, or antibodies to PRASP, and mimetics, agonists,
CC   antagonists or inhibitors of PRASP, are used for treating or preventing
CC   a reproductive disorder or cancer. Examples of reproductive disorder
CC   include, abnormal prolactin production, infertility, tubal disease,
CC   ovulatory defects, endometriosis, polycystic ovary syndrome, autoimmune
CC   disorders, ectopic pregnancy, breast cancer, abnormal spermatogenesis
CC   and testicular cancer. Examples of cancers which may be treated or
CC   prevented include adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma,
CC   sarcoma, teratocarcinoma, and cancers of the adrenal gland, bladder,
CC   bone, bone marrow, brain, breast, cervix, penis, prostate, salivary
CC   glands, skin, spleen, testis, thymus, thyroid and uterus. A vector
CC   capable of expressing PRASP or an agonist which modulates the activity of
CC   PRASP may be administered to treat or prevent a reproductive disorder or
CC   cancer.
XX
SQ   Sequence 1314 BP; 320 A; 400 C; 337 G; 257 T; 0 other;


  Query Match           0.3%;  Score 18;  DB 21;  Length 1314;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1281 aatgaaaaaaaaaaaaaa 1298


RESULT 197
A26298
ID   A26298 standard; cDNA; 1350 BP.
XX
AC   A26298;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 18 SEQ ID NO:28.
XX
KW   Human; secreted protein; diagnosis; neuroprotective; nootropic;
KW   neuroleptic; antimanic; cerebroprotective; immunomodulatory;
KW   anti-microbial; cardiant; cytostatic; antiinflammatory; haemostatic;
KW   anticonvulsant; vasotropic; vaccine; gene therapy; anti-sense therapy;
KW   neural; reproductive; immune disorder; immunodeficiency; infection;
KW   lymphoma; demyelinating disease; autoimmunity; cancer; inflammation;
```

```
KW   aneurysm; haemorrhage; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; Tourette syndrome; multiple sclerosis; meningitis;
KW   ischaemia; mania; dementia; obsessive compulsive disorder;
KW   viral prophylaxis; developmental disorder; sexually-linked disorder;
KW   cardiovascular disorder; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200011014-A1.
XX
PD   02-MAR-2000.
XX
PF   24-AUG-1999;   99WO-US19330.
XX
PR   25-AUG-1998;   98US-0097917.
PR   31-AUG-1998;   98US-0098634.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Moore PA,  Ruben SM,  Olsen HS,  Shi Y,  Rosen CA,  Florence KA;
PI   Soppet DR,  Lafleur DW,  Endress GA,  Ebner R,  Komatsoulis G;
PI   Duan RD;
XX
DR   WPI; 2000-224656/19.
DR   P-PSDB; Y91363.
XX
PT   Novel secreted proteins and corresponding DNA molecules that can be
PT   used to prevent, treat and diagnose disease in humans, for example,
PT   Alzheimer's, cancer, and immune disorders -
XX
PS   Claim 1; Page 322-323; 416pp; English.
XX
CC   The polynucleotide sequences given in A26281 to A26336 encode the human
CC   secreted proteins given in Y91346 to Y91449. The human secreted proteins
CC   can have activities based on the tissues and cells they are expressed in.
CC   Examples of the activities are: neuroprotective; nootropic; neuroleptic;
CC   antimanic; cerebroprotective; immunomodulatory; anti-microbial; cardiant;
CC   cytostatic; antiinflammatory; haemostatic; anticonvulsant; and
CC   vasotropic. The polynucleotides and proteins may be used to prevent,
CC   treat or ameliorate a medical condition, e.g. by protein or gene therapy.
CC   Conditions treatable by the proteins of the invention include neural,
CC   reproductive, or immune disorders, especially immunodeficiency,
CC   infection, lymphomas, demyelinating diseases, auto-immunities, cancer,
CC   general microbial infection, inflammation, aneurysms and haemorrhages.
CC   Specific examples include: Alzheimer's disease; Parkinson's;
CC   Huntington's; Tourette syndrome; multiple sclerosis; meningitis;
CC   ischaemia; prostate cancer; mania; dementia; obsessive compulsive
CC   disorder and viral prophylaxis. The polynucleotides and proteins can also
CC   be used in the detection of disorders associated with the function of the
CC   protein, for example, the detection of developmental disorders,
CC   sexually-linked disorders, or disorders of the cardiovascular system.
```



```
CC   They may also be used as food additives or preservatives. A26272 to
CC   A26280 and Y91345 are sequences used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 1350 BP; 458 A; 205 C; 196 G; 491 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 21;  Length 1350;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1327 aatgaaaaaaaaaaaaaa 1344


RESULT 198
Q11505/c
ID   Q11505 standard; DNA; 1357 BP.
XX
AC   Q11505;
XX
DT   02-JUL-1991  (first entry)
XX
DE   Yeast promoter for glycerol-3-phosphate dehydrogenase.
XX
KW   Transcription control; regulation; fermentation; ss.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FT   promoter        42..52
FT                   /*tag=  a
FT                   /function= RNA POLIII promoter box A
FT   promoter        86..96
FT                   /*tag=  b
FT                   /function= RNA POLIII promoter box B
FT   promoter        113..118
FT                   /*tag=  c
FT                   /function= RNA POLIII terminator
FT   protein_bind    1091..1103
FT                   /*tag=  d
FT                   /bound_moiety= RAPI/GRFI/TUPI
FT   protein_bind    1106..1118
FT                   /*tag=  e
FT                   /bound_moiety= RAPI/GRFI/TUPI
FT   misc_signal     1176..1241
FT                   /*tag=  f
FT                   /function= Pyrimidine (CT) block
FT   TATA_signal     1326..1335
FT                   /*tag=  g
```



```
FT   misc_signal     1369..1406
FT                   /*tag=  h
FT                   /function= Pyrimidine (CT) block
FT   misc_signal     1418..1421
FT                   /*tag=  i
FT                   /function= CAAG box
FT   misc_signal     1425..1429
FT                   /*tag=  j
FT                   /function= CCAAT box
XX
PN   EP424117-A.
XX
PD   24-APR-1991.
XX
PF   17-OCT-1990;   90EP-0311381.
XX
PR   18-OCT-1989;   89GB-0023521.
PR   17-OCT-1990;   90GB-0022507.
XX
PA   (DELT-) DELTA BIOTECHN LTD.
XX
PI   Goodey AR,  Sleep D,  Vakeria D;
XX
DR   WPI; 1991-119346/17.
XX
PT   Yeast promoter for glycerol-3-phosphate dehydrogenase - used to
PT   regulate expression of heterologous genes in yeast.
XX
PS   Claim 1; Page 12; 35pp; English.
XX
CC   The GUT2 promoter was obtd. from a genomic library of fragments
CC   obtd. by BglII restriction of yeast. The promoter may be used to
CC   contruct expression plasmids for the prodn. of heterologous
CC   proteins in yeast. Expression can be regulated by the addn. of
CC   glycerol or ethanol to the fermentation medium. See also Q11506.
XX
SQ   Sequence 1357 BP; 380 A; 358 C; 245 G; 374 T; 0 other;


  Query Match             0.3%;  Score 18;  DB 12;  Length 1357;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     387 tgaaaaaaaaaaaaaatg 404
           ||||||||||||||||||
Db     436 TGAAAAAAAAAAAAAATG 419


RESULT 199
V38741
ID   V38741 standard; cDNA; 1358 BP.
```



```
XX
AC   V38741;
XX
DT   18-SEP-1998  (first entry)
XX
DE   cDNA nfSPI2-1358 encoding a flea serine protease inhibitor protein.
XX
KW   Serine protease inhibitor; SPI; anti-haematophagous ectoparasite;
KW   vaccine; control; flea; acarina; immune response; ss.
XX
OS   Ctenocephalides felis.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..1201
FT                   /*tag=  a
FT                   /product= SPI
XX
PN   WO9820034-A2.
XX
PD   14-MAY-1998.
XX
PF   05-NOV-1997;   97WO-US20678.
XX
PR   07-NOV-1996;   96US-0745995.
XX
PA   (HESK-) HESKA CORP.
XX
PI   Brandt KS,  Maddux JD,  Silver GM,  Wisnewski N;
XX
DR   WPI; 1998-286864/25.
DR   P-PSDB; W62629.
XX
PT   New nucleic acid encoding new serine protease inhibitors from fleas
PT   - and related antibodies, inhibitors, recombinant viruses and cells,
PT   useful for control of fleas on animals, especially cats and dogs
XX
PS   Claim 2; Pages 101-103; 176pp; English.
XX
CC   The present sequence, designated nfSPI2-1358, encodes a Ctenocephalides
CC   felis serine protease inhibitor (SPI) protein. The complementary strand
CC   is represented by V38742. The SPI protein represents a novel target for
CC   anti-haematophagous ectoparasitic vaccines and drugs. Compositions
CC   containing the inhibitor are used to control fleas (adults or larvae) on
CC   mammals or birds, particularly cats and dogs. More generally the
CC   composition is also used for control of insects and acarina, on animals
CC   or on stored goods, plants, trees etc. including domestic pests and
CC   insects that are vectors of disease. The SPI proteins are able to induce
CC   a protective immune response against SPI.
XX
SQ   Sequence 1358 BP; 436 A; 248 C; 263 G; 411 T; 0 other;
```

Query Match 0.3%; Score 18; DB 19; Length 1358;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 816 caacaaagacaatctcat 833
||||||||||||||||||
Db 139 caacaaagacaatctcat 156

RESULT 200
V38742/c
ID V38742 standard; cDNA; 1358 BP.
XX
AC V38742;
XX
DT 18-SEP-1998 (first entry)
XX
DE Complementary strand of cDNA nfSPI2-1358 encoding a SPI protein.
XX
KW Serine protease inhibitor; SPI; anti-haematophagous ectoparasite;
KW vaccine; control; flea; acarina; immune response; ss.
XX
OS Ctenocephalides felis.
XX
PN WO9820034-A2.
XX
PD 14-MAY-1998.
XX
PF 05-NOV-1997; 97WO-US20678.
XX
PR 07-NOV-1996; 96US-0745995.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Maddux JD, Silver GM, Wisnewski N;
XX
DR WPI; 1998-286864/25.
XX
PT New nucleic acid encoding new serine protease inhibitors from fleas
PT - and related antibodies, inhibitors, recombinant viruses and cells,
PT useful for control of fleas on animals, especially cats and dogs
XX
PS Claim 2; Pages 104-105; 176pp; English.
XX
CC The present sequence represents the complementary strand of nfSPI2-1358,
CC which encodes a Ctenocephalides felis serine protease inhibitor (SPI)
CC protein. The SPI protein represents a novel target for
CC anti-haematophagous ectoparasitic vaccines and drugs. Compositions
CC containing the inhibitor are used to control fleas (adults or larvae) on
CC mammals or birds, particularly cats and dogs. More generally the

CC composition is also used for control of insects and acarina, on animals
CC or on stored goods, plants, trees etc. including domestic pests and
CC insects that are vectors of disease. The SPI proteins are able to induce
CC a protective immune response against SPI.
XX
SQ Sequence 1358 BP; 411 A; 263 C; 248 G; 436 T; 0 other;

Query Match 0.3%; Score 18; DB 19; Length 1358;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 816 caacaaagacaatctcat 833
||||||||||||||||||
Db 1220 CAACAAAGACAATCTCAT 1203

Search completed: March 26, 2001, 22:47:15
Job time: 28931 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 12:15:51 ; Search time 6804.77 Seconds
(without alignments)
6101.509 Million cell updates/sec

Title: US-09-410-835-4
Perfect score: 5925
Sequence: 1 ctccattttaagcaactcca.....ttttagaatagacaaagctt 5925

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 7991742 seqs, 3503743858 residues

Word size : 15

Total number of hits satisfying chosen parameters: 272246

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*

22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*
44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*

73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*
117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estro14:*
121: em_estro15:*
122: em_estro16:*
123: em_estro17:*

124: em_estro18:*
125: em_estro19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est75:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: em_gss5:*
159: em_gss6:*
160: gb_gss7:*
161: gb_gss8:*
162: gb_gss9:*
163: em_gss7:*
164: em_gss8:*
165: gb_gss11:*
166: gb_gss10:*
167: em_gss9:*
168: em_gss10:*
169: em_gss11:*
170: em_gss12:*
171: gb_gss12:*
172: gb_gss13:*
173: gb_gss14:*
174: gb_gss15:*

175: gb_gss16:*
176: gb_gss17:*
177: gb_gss18:*
178: gb_gss19:*
179: em_gss13:*
180: gb_gss20:*
181: gb_gss21:*
182: gb_gss22:*
183: gb_gss23:*
184: gb_gss24:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*
190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	23	0.4	621	110	BE577445	BE577445 L48-2287T
c 2	23	0.4	646	110	BE577469	BE577469 L48-2315T
c 3	22	0.4	260	131	BB464908	BB464908 BB464908
c 4	22	0.4	277	131	BB485782	BB485782 BB485782
c 5	22	0.4	294	132	BB495792	BB495792 BB495792
6	22	0.4	411	22	AI593494	AI593494 vu51e02.x
7	22	0.4	416	105	BE237300	BE237300 146442 MA
c 8	22	0.4	483	136	BE852924	BE852924 uw36b10.x
c 9	22	0.4	519	93	AW637516	AW637516 b159d06.w
10	22	0.4	539	36	AV385015	AV385015 AV385015
c 11	22	0.4	592	180	AZ384548	AZ384548 1M0142C11
c 12	22	0.4	1026	138	BE963593	BE963593 60165739H
13	21	0.4	150	28	AL386441	AL386441 MtBC34F09
c 14	21	0.4	157	157	AQ444194	AQ444194 GSSTcO245
c 15	21	0.4	172	15	AI054251	AI054251 qi77b11.x
16	21	0.4	176	29	AU092452	AU092452 AU092452
c 17	21	0.4	185	25	AI836709	AI836709 UI-M-AJ0-
18	21	0.4	213	96	AW911034	AW911034 ur77d09.y
c 19	21	0.4	230	25	AI835982	AI835982 UI-M-AQ0-
20	21	0.4	233	135	BE751829	BE751829 204054 MA
c 21	21	0.4	237	34	AV290224	AV290224 AV290224
c 22	21	0.4	247	190	CNS00Q5F	AL085377 Arabidops

c 23	21	0.4	254	14	AA997094	AA997094 UI-R-C0-h
c 24	21	0.4	254	24	AI707033	AI707033 UI-R-AD1-
c 25	21	0.4	262	28	AL047499	AL047499 DKFZp586A
c 26	21	0.4	262	126	BB251234	BB251234 BB251234
27	21	0.4	265	28	AL047498	AL047498 DKFZp586A
c 28	21	0.4	271	91	AW522628	AW522628 UI-R-BO0-
29	21	0.4	289	101	BB133253	BB133253 BB133253
c 30	21	0.4	292	140	C97838	C97838 C97838 Rice
31	21	0.4	314	29	AU108172	AU108172 AU108172
c 32	21	0.4	323	16	AI111452	AI111452 UI-R-Y0-1
c 33	21	0.4	323	24	AI764576	AI764576 UI-R-Y0-a
c 34	21	0.4	330	23	AI698546	AI698546 wa79c05.x
35	21	0.4	366	144	R84108	R84108 16067 Lambd
c 36	21	0.4	372	15	AI060180	AI060180 UI-R-C1-1
c 37	21	0.4	374	39	AW048532	AW048532 UI-M-BH1-
c 38	21	0.4	376	27	AI969768	AI969768 wq73f07.x
c 39	21	0.4	377	23	AI641714	AI641714 fc22h12.x
40	21	0.4	380	105	BE227191	BE227191 us78a05.y
c 41	21	0.4	387	28	AL041690	AL041690 DKFZp434N
c 42	21	0.4	391	91	AW528569	AW528569 UI-R-BT1-
c 43	21	0.4	400	2	AA132821	AA132821 zo18804.r
c 44	21	0.4	400	144	R63336	R63336 yi07f09.r1
c 45	21	0.4	401	16	AI117620	AI117620 ui48b02.x
c 46	21	0.4	407	8	AA505862	AA505862 nh99e04.s
c 47	21	0.4	408	25	AI842076	AI842076 UI-M-AN1-
48	21	0.4	409	28	AL377900	AL377900 MtBB34F10
49	21	0.4	410	140	D76720	D76720 MUS69B04 mo
50	21	0.4	413	138	BE951351	BE951351 UI-M-CC0-
c 51	21	0.4	415	144	R33351	R33351 yh77f10.r1
52	21	0.4	419	150	AQ198451	AQ198451 RPCI11-48
53	21	0.4	423	89	AW321430	AW321430 uo31g02.y
54	21	0.4	428	12	AA855552	AA855552 vw74f12.r
55	21	0.4	430	28	AL377901	AL377901 MtBB34F10
56	21	0.4	431	135	BE756820	BE756820 211079 MA
c 57	21	0.4	436	109	BE508093	BE508093 dc26e02.y
c 58	21	0.4	438	190	CNS00QEZ	AL085721 Arabidops
c 59	21	0.4	439	26	AI911383	AI911383 wd16e02.x
c 60	21	0.4	440	28	AL368910	AL368910 MtBA27F02
c 61	21	0.4	442	20	AI425943	AI425943 mg27d07.x
c 62	21	0.4	448	138	BE954247	BE954247 UI-M-CC1-
c 63	21	0.4	450	149	AQ111049	AQ111049 CIT-HSP-2
c 64	21	0.4	454	24	AI715886	AI715886 UI-R-Y0-a
c 65	21	0.4	454	180	AZ395583	AZ395583 1M0159F18
c 66	21	0.4	457	39	AW074370	AW074370 xb11h04.x
c 67	21	0.4	460	19	AI378562	AI378562 tc72a10.x
c 68	21	0.4	461	137	BE943917	BE943917 UI-M-BH3-
c 69	21	0.4	463	15	AI071557	AI071557 UI-R-C1-k
c 70	21	0.4	464	88	AW297911	AW297911 UI-H-BW0-
c 71	21	0.4	470	8	AA533164	AA533164 nj60f12.s
c 72	21	0.4	479	134	BE099767	BE099767 UI-R-BJ1-
c 73	21	0.4	480	14	AA964331	AA964331 UI-R-C0-g

c 74	21	0.4	492	28	AT002267	AT002267 AT002267
75	21	0.4	497	28	AL374239	AL374239 MtBB05E01
76	21	0.4	500	28	AL374022	AL374022 MtBB04C01
77	21	0.4	509	136	BE809028	BE809028 214367 MA
78	21	0.4	511	150	AQ218828	AQ218828 HS_3248_B
79	21	0.4	513	28	AL385814	AL385814 MtBC30G12
80	21	0.4	519	29	AU096044	AU096044 AU096044
c 81	21	0.4	522	12	AA832599	AA832599 vw43f03.r
82	21	0.4	533	134	BE033628	BE033628 MF05C02 M
83	21	0.4	542	28	AL378418	AL378418 MtBB38B03
c 84	21	0.4	542	39	AW028928	AW028928 wv97f08.x
85	21	0.4	546	25	AI783160	AI783160 BSBmL3SZ1
86	21	0.4	551	37	AV549932	AV549932 AV549932
87	21	0.4	574	184	B96756	B96756 T32F9TR TAM
c 88	21	0.4	587	156	AQ346762	AQ346762 RPCI11-12
c 89	21	0.4	598	177	AZ235445	AZ235445 RPCI-23-7
90	21	0.4	606	36	AV386003	AV386003 AV386003
91	21	0.4	614	38	AV699504	AV699504 AV699504
c 92	21	0.4	657	13	AA920399	AA920399 vx95g12.r
c 93	21	0.4	669	192	CNS03M8P	AL250450 Tetraodon
c 94	21	0.4	670	160	AQ491144	AQ491144 RPCI-11-2
95	21	0.4	692	180	AZ382945	AZ382945 1M0140J23
96	21	0.4	712	149	AQ110183	AQ110183 CIT-HSP-2
97	21	0.4	722	176	AZ183754	AZ183754 SP_1001_B
98	21	0.4	742	191	CNS0252V	AL181552 Tetraodon
c 99	21	0.4	757	176	AZ195068	AZ195068 SP_1029_A
c 100	21	0.4	781	106	BE269463	BE269463 601184638
c 101	21	0.4	791	174	AZ046289	AZ046289 nbeb0091J
102	21	0.4	955	191	CNS01ZTJ	AL174736 Tetraodon
c 103	20	0.3	65	27	AI963429	AI963429 wt61b05.x
104	20	0.3	76	2	AA110380	AA110380 mp43e03.r
c 105	20	0.3	82	26	AI913476	AI913476 tz77h02.x
106	20	0.3	109	7	AA399932	AA399932 vd67f07.r
c 107	20	0.3	120	89	AW333457	AW333457 S22A9 AGS
c 108	20	0.3	122	21	AI537111	AI537111 to15h03.x
c 109	20	0.3	128	104	BE104648	BE104648 UI-R-BX0-
c 110	20	0.3	130	21	AI540717	AI540717 tn72f01.x
c 111	20	0.3	134	92	AW576296	AW576296 UI-HF-BN0
c 112	20	0.3	148	90	AW444585	AW444585 UI-H-BI3-
c 113	20	0.3	160	40	AW138945	AW138945 UI-H-BI1-
c 114	20	0.3	167	102	BB170597	BB170597 BB170597
c 115	20	0.3	177	36	AV378192	AV378192 AV378192
c 116	20	0.3	190	23	AI688140	AI688140 wa83g10.x
117	20	0.3	192	136	BE808167	BE808167 213308 MA
c 118	20	0.3	194	137	BE946267	BE946267 UI-M-BH3-
c 119	20	0.3	201	15	AI059741	AI059741 UI-R-C1-1
120	20	0.3	201	89	AW357010	AW357010 39751 MAR
121	20	0.3	202	142	H63232	H63232 yr51a02.s1
c 122	20	0.3	205	18	AI247269	AI247269 qj60f09.x
123	20	0.3	206	140	C84740	C84740 C84740 Dict
c 124	20	0.3	207	13	AA903342	AA903342 ok45h10.s

c 125	20	0.3	210	138	BE995735	BE995735 UI-M-CG0p
c 126	20	0.3	215	145	T10769	T10769 hbc1239 Hum
c 127	20	0.3	221	99	BB077862	BB077862 BB077862
c 128	20	0.3	225	35	AV346494	AV346494 AV346494
129	20	0.3	227	21	AI544518	AI544518 fb76b11.x
c 130	20	0.3	227	39	AW050394	AW050394 wz11f05.x
c 131	20	0.3	229	35	AV354500	AV354500 AV354500
c 132	20	0.3	230	99	BB077631	BB077631 BB077631
c 133	20	0.3	233	138	BE991125	BE991125 UI-M-BZ1-
c 134	20	0.3	233	138	BE991210	BE991210 UI-M-BZ1-
c 135	20	0.3	238	24	AI764202	AI764202 UI-R-Y0-a
136	20	0.3	238	39	AW070146	AW070146 mgie1005D
137	20	0.3	238	136	BE804542	BE804542 sr83a04.y
c 138	20	0.3	240	18	AI263668	AI263668 qx67f09.x
139	20	0.3	240	29	AU073502	AU073502 AU073502
140	20	0.3	240	29	AU074261	AU074261 AU074261
c 141	20	0.3	242	147	Z21539	Z21539 HSAFIB007 S
c 142	20	0.3	244	22	AI623688	AI623688 ts22h10.x
c 143	20	0.3	244	29	AV021342	AV021342 AV021342
144	20	0.3	247	27	AI974381	AI974381 T110323e
145	20	0.3	252	9	AA579530	AA579530 nf37h06.s
c 146	20	0.3	253	40	AW109600	AW109600 gate0004N
147	20	0.3	254	28	AL368864	AL368864 MtBA27C04
148	20	0.3	260	29	AU058025	AU058025 AU058025
c 149	20	0.3	261	12	AA788980	AA788980 ah29b12.s
c 150	20	0.3	262	21	AI481145	AI481145 vg86e04.x
c 151	20	0.3	262	91	AW490767	AW490767 UI-M-BH3-
c 152	20	0.3	262	134	BE052081	BE052081 GA__Ea000
c 153	20	0.3	265	18	AI250156	AI250156 qx49a05.x
c 154	20	0.3	273	138	BE953755	BE953755 UI-M-CC1-
c 155	20	0.3	277	22	AI561501	AI561501 vw88c12.x
c 156	20	0.3	283	173	AQ976175	AQ976175 RPCI-23-3
c 157	20	0.3	284	128	BB366530	BB366530 BB366530
158	20	0.3	285	29	AU036991	AU036991 AU036991
159	20	0.3	285	126	BB265787	BB265787 BB265787
160	20	0.3	287	126	BB264068	BB264068 BB264068
161	20	0.3	291	103	BB246243	BB246243 BB246243
c 162	20	0.3	292	25	AI835732	AI835732 UI-M-AI0-
c 163	20	0.3	294	22	AI602026	AI602026 UI-R-Y0-u
c 164	20	0.3	294	23	AI661433	AI661433 va26e12.x
c 165	20	0.3	294	91	AW492508	AW492508 UI-M-BH3-
c 166	20	0.3	294	104	BE116692	BE116692 UI-R-BS1-
167	20	0.3	295	146	W10380	W10380 ma38c07.r1
c 168	20	0.3	298	16	AI084635	AI084635 oz78d04.x
169	20	0.3	298	29	AU077610	AU077610 AU077610
170	20	0.3	299	29	AU057893	AU057893 AU057893
c 171	20	0.3	299	38	AW016180	AW016180 UI-H-BI0p
172	20	0.3	299	104	BE108199	BE108199 UI-R-BS1-
173	20	0.3	301	183	B05848	B05848 cSRL-72a5-u
c 174	20	0.3	303	25	AI842938	AI842938 UI-M-A01-
c 175	20	0.3	305	10	AA669251	AA669251 ac17e11.s

c 176	20	0.3	305	144	R46662	R46662 yg53h05.s1
c 177	20	0.3	305	150	AQ230212	AQ230212 HS_2041_B
c 178	20	0.3	306	138	BE954589	BE954589 UI-M-CC1-
c 179	20	0.3	310	39	AW028947	AW028947 wv91h10.x
c 180	20	0.3	310	138	BF012355	BF012355 ux55b02.y
181	20	0.3	311	138	BE995963	BE995963 UI-M-BZ1-
c 182	20	0.3	312	15	AI073068	AI073068 UI-R-Y0-1
c 183	20	0.3	314	95	AW768811	AW768811 hl53f09.x
184	20	0.3	314	132	BB499835	BB499835 BB499835
c 185	20	0.3	315	88	AW265414	AW265414 xq52d12.x
186	20	0.3	315	126	BB276510	BB276510 BB276510
c 187	20	0.3	318	23	AI682836	AI682836 wc66h12.x
c 188	20	0.3	318	98	BB021002	BB021002 BB021002
c 189	20	0.3	319	20	AI471617	AI471617 ti66a11.x
c 190	20	0.3	319	24	AI753085	AI753085 cr04h02.x
c 191	20	0.3	319	161	AQ567887	AQ567887 HS_2117_A
c 192	20	0.3	320	90	AW457128	AW457128 UI-M-BH3-
c 193	20	0.3	320	138	BE995761	BE995761 UI-M-BZ1-
194	20	0.3	321	148	AQ071286	AQ071286 HS_2266_B
195	20	0.3	324	18	AI248370	AI248370 qh76b01.x
196	20	0.3	324	93	AW619259	AW619259 619 MARC
c 197	20	0.3	327	91	AW528408	AW528408 UI-R-BO1-
c 198	20	0.3	328	15	AI070265	AI070265 UI-R-Y0-1
199	20	0.3	330	28	AL047858	AL047858 DKFZp586L
200	20	0.3	330	141	H42282	H42282 yo63d03.r1
201	20	0.3	330	149	AQ104978	AQ104978 HS_2168_B
c 202	20	0.3	331	11	AA737046	AA737046 nx90h12.s
203	20	0.3	331	28	AL389538	AL389538 MtBC55E10
c 204	20	0.3	333	87	AW193105	AW193105 xl70c11.x
c 205	20	0.3	334	11	AA760127	AA760127 vv73d10.r
206	20	0.3	335	174	AZ044701	AZ044701 Gm_UMb001
c 207	20	0.3	342	38	AW014055	AW014055 UI-H-BI0-
c 208	20	0.3	343	25	AI845447	AI845447 UI-M-AO1-
209	20	0.3	344	91	AW492575	AW492575 UI-M-BH3-
c 210	20	0.3	348	14	AA955671	AA955671 UI-R-E1-f
211	20	0.3	348	143	N98132	N98132 2227C3 czap
c 212	20	0.3	349	12	AA806740	AA806740 ob91a07.s
c 213	20	0.3	350	22	AI619304	AI619304 486091D08
c 214	20	0.3	350	104	BE121052	BE121052 UI-R-CA0-
c 215	20	0.3	353	12	AA810521	AA810521 ob62b06.s
216	20	0.3	353	135	BE750810	BE750810 202448 MA
c 217	20	0.3	354	13	AA914162	AA914162 vy98b06.r
218	20	0.3	354	157	AQ444192	AQ444192 GSSTc0246
c 219	20	0.3	355	18	AI300191	AI300191 qn60d08.x
c 220	20	0.3	355	23	AI676759	AI676759 TENG01072
221	20	0.3	355	25	AI797160	AI797160 we27f06.x
c 222	20	0.3	360	139	C60419	C60419 C60419 Yuji
c 223	20	0.3	361	12	AA811305	AA811305 ob76c02.s
c 224	20	0.3	362	91	AW524925	AW524925 UI-R-BO0-
c 225	20	0.3	364	1	AA018793	AA018793 ze58a06.r
226	20	0.3	365	29	AU061026	AU061026 AU061026

c 227	20	0.3	365	88	AW294271	AW294271 UI-H-BI2-
228	20	0.3	367	90	AW432681	AW432681 sh85a04.y
229	20	0.3	368	89	AW365523	AW365523 MR0-HT007
230	20	0.3	368	162	AQ663242	AQ663242 HS_5463_B
c 231	20	0.3	369	19	AI395322	AI395322 MA002913.
c 232	20	0.3	372	91	AW514505	AW514505 xu63e12.x
233	20	0.3	372	107	BE411603	BE411603 ISC007.C1
c 234	20	0.3	372	138	BE996126	BE996126 UI-M-BZ1-
c 235	20	0.3	373	160	AQ508997	AQ508997 RPCI-11-2
c 236	20	0.3	379	15	AI029163	AI029163 UI-R-C0-i
237	20	0.3	379	135	BE752083	BE752083 204385 MA
c 238	20	0.3	381	15	AI050086	AI050086 an35d01.x
c 239	20	0.3	381	37	AV530508	AV530508 AV530508
c 240	20	0.3	382	90	AW440091	AW440091 xs81h09.x
c 241	20	0.3	384	92	AW576558	AW576558 UI-HF-BN0
c 242	20	0.3	385	12	AA827711	AA827711 ob54e10.s
c 243	20	0.3	385	15	AI060233	AI060233 UI-R-C1-i
c 244	20	0.3	385	90	AW452830	AW452830 UI-H-BI3-
c 245	20	0.3	385	104	BE116114	BE116114 UI-R-BS1-
c 246	20	0.3	389	25	AI847212	AI847212 UI-M-AI1-
c 247	20	0.3	389	90	AW455833	AW455833 UI-M-BH3-
c 248	20	0.3	390	24	AI709443	AI709443 UI-R-C3-s
249	20	0.3	392	28	AL367214	AL367214 MtBA12H08
250	20	0.3	392	89	AW365517	AW365517 MR0-HT007
251	20	0.3	392	104	BE141205	BE141205 MR0-HT007
c 252	20	0.3	393	38	AW014287	AW014287 UI-H-BI0-
c 253	20	0.3	393	91	AW530726	AW530726 UI-R-BT1-
c 254	20	0.3	394	91	AW529292	AW529292 UI-R-BT1-
c 255	20	0.3	395	90	AW449935	AW449935 UI-H-BI3-
c 256	20	0.3	396	38	AW016433	AW016433 UI-H-BI0p
c 257	20	0.3	396	147	W54258	W54258 md04c12.r1
258	20	0.3	397	135	BE750805	BE750805 202443 MA
c 259	20	0.3	399	15	AI027226	AI027226 ov83d05.x
260	20	0.3	399	89	AW365507	AW365507 MR0-HT007
c 261	20	0.3	400	13	AA875404	AA875404 UI-R-E0-c
262	20	0.3	401	89	AW365510	AW365510 MR0-HT007
c 263	20	0.3	401	144	R38496	R38496 yf60e01.s1
c 264	20	0.3	403	15	AI071452	AI071452 UI-R-C1-k
265	20	0.3	403	134	BE047612	BE047612 tz39d07.y
c 266	20	0.3	404	21	AI537195	AI537195 tp01d09.x
267	20	0.3	404	93	AW643607	AW643607 cm31g01.w
268	20	0.3	404	106	BE334427	BE334427 us79a05.y
c 269	20	0.3	405	18	AI248461	AI248461 qh71b09.x
c 270	20	0.3	405	92	AW576522	AW576522 UI-HF-BN0
c 271	20	0.3	406	13	AA933024	AA933024 oo42h05.s
c 272	20	0.3	406	19	AI395321	AI395321 MA002911.
273	20	0.3	409	147	W59088	W59088 md74c01.r1
c 274	20	0.3	410	157	AQ416192	AQ416192 RPCI-11-1
c 275	20	0.3	411	17	AI214205	AI214205 qp40f12.x
276	20	0.3	411	89	AW365505	AW365505 MR0-HT007
c 277	20	0.3	413	24	AI745668	AI745668 tr24f01.x

278	20	0.3	413	89	AW365509	AW365509 MR0-HT007
c 279	20	0.3	414	19	AI343300	AI343300 tb94f10.x
c 280	20	0.3	415	23	AI681468	AI681468 tx37f07.x
c 281	20	0.3	415	141	H18831	H18831 ym45c10.s1
c 282	20	0.3	416	18	AI298413	AI298413 qn01f08.x
c 283	20	0.3	417	40	AW150717	AW150717 xg38c03.x
284	20	0.3	418	89	AW365496	AW365496 MR0-HT007
285	20	0.3	418	104	BE141214	BE141214 MR0-HT007
c 286	20	0.3	420	16	AI140800	AI140800 oo15b02.x
c 287	20	0.3	421	10	AA636058	AA636058 nr71g04.s
c 288	20	0.3	421	27	AI972692	AI972692 wr42b06.x
c 289	20	0.3	422	16	AI140606	AI140606 qe05f04.x
290	20	0.3	422	22	AI593545	AI593545 vu55d05.x
c 291	20	0.3	422	27	AI954152	AI954152 wx80d05.x
c 292	20	0.3	422	39	AW074699	AW074699 xa90e06.x
293	20	0.3	423	89	AW365499	AW365499 MR0-HT007
c 294	20	0.3	424	39	AW047199	AW047199 UI-M-BH1-
c 295	20	0.3	424	138	BE981855	BE981855 UI-M-CG0p
c 296	20	0.3	425	19	AI375862	AI375862 tc13g01.x
c 297	20	0.3	425	144	R81613	R81613 yj04f07.r1
c 298	20	0.3	427	25	AI792460	AI792460 an34e08.y
c 299	20	0.3	427	27	AI970192	AI970192 wr08d04.x
c 300	20	0.3	428	11	AA731326	AA731326 nw58a03.s
c 301	20	0.3	428	12	AA825489	AA825489 od04d07.s
c 302	20	0.3	428	17	AI200842	AI200842 qf67h02.x
c 303	20	0.3	428	138	BF000002	BF000002 7h15f09.x
304	20	0.3	431	89	AW325086	AW325086 TENU4391
c 305	20	0.3	433	17	AI200972	AI200972 qf68h01.x
306	20	0.3	433	97	AW964548	AW964548 EST376621
c 307	20	0.3	434	13	AA901059	AA901059 UI-R-A1-d
c 308	20	0.3	434	88	AW281863	AW281863 fj58c12.x
309	20	0.3	434	104	BE141210	BE141210 MR0-HT007
c 310	20	0.3	435	134	BE026666	BE026666 db29g05.x
c 311	20	0.3	437	2	AA117434	AA117434 mn21g11.r
c 312	20	0.3	438	11	AA760736	AA760736 nz10a11.s
c 313	20	0.3	439	104	BE117900	BE117900 UI-R-BS1-
c 314	20	0.3	441	94	AW744503	AW744503 ur33h12.y
c 315	20	0.3	443	16	AI094675	AI094675 oy62g08.s
c 316	20	0.3	444	144	R51884	R51884 yg78c01.s1
317	20	0.3	446	104	BE141224	BE141224 MR0-HT007
c 318	20	0.3	448	14	AA937977	AA937977 oc08f04.s
c 319	20	0.3	448	24	AI716647	AI716647 UI-R-Y0-a
c 320	20	0.3	448	135	BE760495	BE760495 an_1770 A
c 321	20	0.3	449	111	BE691401	BE691401 ut16d05.x
322	20	0.3	450	19	AI402241	AI402241 GH11023.3
c 323	20	0.3	451	22	AI626625	AI626625 fc05f07.x
c 324	20	0.3	452	23	AI662230	AI662230 mz98b09.x
c 325	20	0.3	452	150	AQ230210	AQ230210 HS_2041_B
326	20	0.3	452	183	AZ508615	AZ508615 1M0351A11
c 327	20	0.3	454	28	AL036436	AL036436 DKFZp564C
c 328	20	0.3	455	19	AI395037	AI395037 MA002262.

c 329	20	0.3	456	40	AW135740	AW135740 UI-H-BI1-
c 330	20	0.3	458	22	AI566762	AI566762 tn51h09.x
c 331	20	0.3	459	23	AI678864	AI678864 tu70b08.x
332	20	0.3	459	184	B93750	B93750 CIT-HSP-216
c 333	20	0.3	461	38	AW014146	AW014146 UI-H-BI0-
c 334	20	0.3	462	15	AI044529	AI044529 UI-R-C1-k
c 335	20	0.3	462	137	BE944834	BE944834 UI-M-BH3-
c 336	20	0.3	463	90	AW457644	AW457644 UI-M-BH3-
c 337	20	0.3	464	27	AJ280620	AJ280620 4A3A-AAR-
338	20	0.3	465	29	AU093063	AU093063 AU093063
339	20	0.3	465	89	AW365497	AW365497 MR0-HT007
c 340	20	0.3	466	91	AW531809	AW531809 UI-R-C4-a
c 341	20	0.3	467	22	AI571900	AI571900 tr71a02.x
c 342	20	0.3	468	19	AI395327	AI395327 MA002927.
343	20	0.3	468	147	W47835	W47835 mc82c09.r1
c 344	20	0.3	469	18	AI282923	AI282923 qt69c10.x
c 345	20	0.3	469	97	AW943050	AW943050 LD10826.3
c 346	20	0.3	469	137	BE943679	BE943679 UI-M-BH3-
c 347	20	0.3	470	24	AI755131	AI755131 cr35h01.x
c 348	20	0.3	472	12	AA831337	AA831337 oc74a08.s
c 349	20	0.3	473	27	AI985096	AI985096 wr97c03.x
c 350	20	0.3	475	11	AA743630	AA743630 ny24e11.s
c 351	20	0.3	475	23	AI658890	AI658890 tt98h11.x
c 352	20	0.3	476	13	AA888876	AA888876 ak23f10.s
c 353	20	0.3	479	24	AI754598	AI754598 cr27d11.x
c 354	20	0.3	480	137	BE910743	BE910743 601502185
c 355	20	0.3	482	25	AI830674	AI830674 wj41b04.x
c 356	20	0.3	484	38	AW010706	AW010706 ST11C01 P
357	20	0.3	486	27	AI965273	AI965273 fc88g12.x
c 358	20	0.3	488	110	BE576856	BE576856 LO-1663T3
c 359	20	0.3	489	14	AA964492	AA964492 UI-R-E1-g
c 360	20	0.3	491	40	AW121873	AW121873 UI-M-BH2.
c 361	20	0.3	491	87	AW206273	AW206273 UI-H-BI1-
c 362	20	0.3	495	9	AA598890	AA598890 ae37d10.s
363	20	0.3	496	171	AQ873389	AQ873389 V65A1 mTn
364	20	0.3	497	17	AI176276	AI176276 EST219858
c 365	20	0.3	498	11	AA744802	AA744802 ny26h02.s
c 366	20	0.3	498	25	AI835050	AI835050 UI-M-AI1-
c 367	20	0.3	498	28	AL366464	AL366464 MtBA08B02
368	20	0.3	500	11	AA711937	AA711937 vu59h03.r
c 369	20	0.3	501	138	BE949065	BE949065 UI-M-BH3-
370	20	0.3	501	173	AZ007591	AZ007591 RPCI-23-3
c 371	20	0.3	504	104	BE107730	BE107730 UI-R-CA0-
372	20	0.3	506	28	AL373344	AL373344 MtBA57C05
373	20	0.3	507	172	AQ924713	AQ924713 RPCI-23-2
c 374	20	0.3	508	7	AA410910	AA410910 sv39d06.r
c 375	20	0.3	508	20	AI432649	AI432649 ti22a11.x
c 376	20	0.3	508	20	AI439811	AI439811 ti56d03.x
c 377	20	0.3	514	8	AA535805	AA535805 nf93g04.s
c 378	20	0.3	515	20	AI475354	AI475354 ti82c03.x
379	20	0.3	516	91	AW506291	AW506291 GE1677 G1

c 380	20	0.3	517	177	AZ275931	AZ275931 RPCI-23-1
c 381	20	0.3	518	87	AW204203	AW204203 UI-H-BI1-
c 382	20	0.3	519	39	AW049680	AW049680 UI-M-BH1-
c 383	20	0.3	520	26	AI921297	AI921297 wo23a06.x
384	20	0.3	520	150	AQ241929	AQ241929 CITBI-E1-
c 385	20	0.3	522	10	AA643223	AA643223 nr96h08.s
c 386	20	0.3	522	17	AI188551	AI188551 qd09g11.x
c 387	20	0.3	522	24	AI763037	AI763037 wi64d09.x
c 388	20	0.3	527	21	AI491786	AI491786 to01b01.x
c 389	20	0.3	529	39	AW049279	AW049279 UI-M-BH1-
c 390	20	0.3	529	134	BE098719	BE098719 UI-R-BJ1-
c 391	20	0.3	529	175	AZ109933	AZ109933 RPCI-23-8
c 392	20	0.3	530	90	AW449074	AW449074 UI-H-BI3-
393	20	0.3	535	28	AL379422	AL379422 MtBB45C03
c 394	20	0.3	535	38	AW014587	AW014587 UI-H-BI0p
c 395	20	0.3	536	39	AW069793	AW069793 cr50f09.x
c 396	20	0.3	538	160	AQ484994	AQ484994 RPCI-11-2
c 397	20	0.3	539	14	AA998140	AA998140 UI-R-C0-i
c 398	20	0.3	543	25	AI807137	AI807137 wf25a11.x
c 399	20	0.3	545	138	BE992484	BE992484 UI-M-BZ1-
c 400	20	0.3	546	110	BE627402	BE627402 ut93e08.x
c 401	20	0.3	546	161	AQ570314	AQ570314 HS_5352_B
402	20	0.3	553	160	AQ469421	AQ469421 CITBI-E1-
c 403	20	0.3	554	107	BE362182	BE362182 DG1_84_G1
c 404	20	0.3	556	9	AA596495	AA596495 vm58a06.r
c 405	20	0.3	556	139	BF041852	BF041852 BP250024B
406	20	0.3	558	137	BE875368	BE875368 60148908l
407	20	0.3	559	2	AA141268	AA141268 CK01488.3
c 408	20	0.3	560	174	AZ020795	AZ020795 RPCI-23-3
c 409	20	0.3	561	137	BE945360	BE945360 UI-M-BZ0-
c 410	20	0.3	562	92	AW549870	AW549870 L0056D11-
411	20	0.3	562	134	BE055891	BE055891 GA__Ea002
412	20	0.3	565	2	AA080196	AA080196 mm33e02.r
c 413	20	0.3	569	91	AW488142	AW488142 UI-M-BH3-
c 414	20	0.3	573	23	AI654308	AI654308 tq90b05.x
c 415	20	0.3	577	161	AQ600528	AQ600528 HS_5317_A
c 416	20	0.3	579	18	AI309330	AI309330 qo80d04.x
c 417	20	0.3	579	157	AQ418988	AQ418988 RPCI-11-1
c 418	20	0.3	580	1	AA068786	AA068786 mm60a10.r
419	20	0.3	580	192	CNS04LR2	AL296471 Tetraodon
c 420	20	0.3	582	134	BE100177	BE100177 UI-R-BJ1-
421	20	0.3	585	161	AQ602669	AQ602669 HS_2132_A
c 422	20	0.3	588	89	AW334161	AW334161 S31C5 AGS
c 423	20	0.3	588	182	AZ488395	AZ488395 1M0318H07
c 424	20	0.3	590	26	AI924824	AI924824 wn23d09.x
c 425	20	0.3	595	104	BE113219	BE113219 UI-R-BJ1-
426	20	0.3	602	174	AZ027412	AZ027412 RPCI-23-3
c 427	20	0.3	616	17	AI184603	AI184603 qf46a12.x
428	20	0.3	621	97	AW974548	AW974548 EST386652
429	20	0.3	630	38	AV693666	AV693666 AV693666
c 430	20	0.3	632	89	AW350612	AW350612 GM210009A

431	20	0.3	639	29	AU030787	AU030787 AU030787
432	20	0.3	656	93	AW686088	AW686088 NP038B08N
c 433	20	0.3	657	157	AQ449655	AQ449655 500003A05
434	20	0.3	672	91	AW505991	AW505991 ge1272 G1
435	20	0.3	672	106	BE277717	BE277717 601119918
c 436	20	0.3	688	156	AQ377951	AQ377951 RPCI11-16
437	20	0.3	701	23	AI675810	AI675810 wb97b07.x
438	20	0.3	707	93	AW668225	AW668225 GA__Ea001
439	20	0.3	709	29	AU100718	AU100718 AU100718
c 440	20	0.3	721	26	AI879400	AI879400 au50d03.x
c 441	20	0.3	725	89	AW335324	AW335324 S45H6 AGS
c 442	20	0.3	727	156	AQ330684	AQ330684 nbxb0047B
c 443	20	0.3	731	183	BI22042	BI22042 P20H18-T7 I
c 444	20	0.3	732	15	AI031459	AI031459 30G12-1P
c 445	20	0.3	742	95	AW776535	AW776535 EST335600
c 446	20	0.3	744	138	BE951272	BE951272 UI-M-CC0-
c 447	20	0.3	752	26	AI927967	AI927967 wo92a06.x
c 448	20	0.3	754	28	AL039478	AL039478 DKFZp434P
c 449	20	0.3	778	39	AW087157	AW087157 xb35b09.x
c 450	20	0.3	791	176	AZ195734	AZ195734 SP_1030_B
c 451	20	0.3	829	176	AZ200356	AZ200356 SP_1002_A
452	20	0.3	839	176	AZ190365	AZ190365 SP_1016_B
c 453	20	0.3	844	190	CNS009VF	AL053982 Drosophil
454	20	0.3	896	193	CNS05H73	AL337224 Tetraodon
c 455	20	0.3	908	193	CNS05HPP	AL337894 Tetraodon
456	20	0.3	918	137	BE903590	BE903590 601677588
c 457	20	0.3	929	183	B07973	B07973 P22K10-T7.1
c 458	20	0.3	944	172	AQ900982	AQ900982 HS_3225_A
c 459	20	0.3	944	191	CNS02UC5	AL214286 Tetraodon
460	20	0.3	982	190	CNS008OV	AL052248 Drosophil
c 461	20	0.3	982	193	CNS05ED4	AL333553 Tetraodon
c 462	20	0.3	999	190	CNS012TN	AL102005 Drosophil
463	20	0.3	1005	190	CNS008NA	AL052090 Drosophil
c 464	20	0.3	1101	190	CNS000W8	AL053679 Drosophil
465	20	0.3	1101	190	CNS001KX	AL061136 Drosophil
c 466	20	0.3	1101	190	CNS00FVG	AL071300 Drosophil
c 467	19	0.3	65	19	AI341410	AI341410 qx91d08.x
468	19	0.3	72	27	AJ283066	AJ283066 4A3A-P4F2
c 469	19	0.3	115	143	N63154	N63154 yz37f12.s1
470	19	0.3	120	37	AV560043	AV560043 AV560043
471	19	0.3	122	9	AA574755	AA574755 vm27c01.r
c 472	19	0.3	122	91	AW488415	AW488415 UI-M-BH3-
c 473	19	0.3	122	137	BE946963	BE946963 UI-M-BH3-
c 474	19	0.3	122	138	BE947665	BE947665 UI-M-BH3-
475	19	0.3	122	141	H41223	H41223 yp64g12.s1
c 476	19	0.3	126	13	AA912792	AA912792 o131c12.s
477	19	0.3	130	29	AU053964	AU053964 AU053964
478	19	0.3	132	145	T62202	T62202 yb97c01.s1
479	19	0.3	144	175	AZ116144	AZ116144 RPCI-23-4
c 480	19	0.3	146	138	BE980486	BE980486 UI-M-BG2-
c 481	19	0.3	159	28	AL366912	AL366912 MtBA11A02

c 482	19	0.3	164	25	AI836277	AI836277 UI-M-AQ0-
c 483	19	0.3	165	13	AA902288	AA902288 ok69g12.s
484	19	0.3	166	11	AA733174	AA733174 sg77h04.s
c 485	19	0.3	166	88	AW252509	AW252509 UI-R-BJ0-
c 486	19	0.3	166	104	BE105364	BE105364 UI-R-BX0-
c 487	19	0.3	168	33	AV262755	AV262755 AV262755
488	19	0.3	175	27	AI933607	AI933607 wn41d09.x
489	19	0.3	178	138	BE954230	BE954230 UI-M-CC1-
490	19	0.3	178	146	W24467	W24467 zb60e04.r1
c 491	19	0.3	181	14	AA957699	AA957699 UI-R-E1-f
c 492	19	0.3	185	22	AI575167	AI575167 UI-R-G0-u
493	19	0.3	185	130	BB438390	BB438390 BB438390
494	19	0.3	186	17	AI174694	AI174694 HA2393 Hu
495	19	0.3	186	88	AW275440	AW275440 xp36f11.x
c 496	19	0.3	186	137	BE946073	BE946073 UI-M-BH3-
c 497	19	0.3	188	26	AI886888	AI886888 wl30b09.x
498	19	0.3	188	129	BB379476	BB379476 BB379476
c 499	19	0.3	189	144	R46805	R46805 yg54c08.s1
c 500	19	0.3	190	4	AA262729	AA262729 zs24a10.s
c 501	19	0.3	190	101	BB148043	BB148043 BB148043
c 502	19	0.3	190	151	AQ260758	AQ260758 CITBI-E1-
c 503	19	0.3	194	23	AI650428	AI650428 wa90f11.x
c 504	19	0.3	194	109	BE503245	BE503245 7a14b06.x
c 505	19	0.3	194	177	AZ236999	AZ236999 RPCI-23-7
c 506	19	0.3	195	104	BE121083	BE121083 UI-R-CA0-
507	19	0.3	201	29	AU061254	AU061254 AU061254
508	19	0.3	201	109	BE521097	BE521097 M16H9STM
509	19	0.3	203	94	AW690262	AW690262 NF028G05S
510	19	0.3	204	28	AL378215	AL378215 MtBB36G04
511	19	0.3	204	138	BE993017	BE993017 UI-M-BZ1-
c 512	19	0.3	206	14	AA937874	AA937874 nw90d11.s
c 513	19	0.3	206	109	BE521098	BE521098 M16H9XP A
514	19	0.3	206	136	BE806939	BE806939 sa08h03.y
515	19	0.3	208	7	AA472363	AA472363 vh06e11.r
c 516	19	0.3	208	28	AL079556	AL079556 DKFZp434H
c 517	19	0.3	209	11	AA772969	AA772969 af33h04.s
518	19	0.3	211	95	AW762979	AW762979 ur59f02.y
c 519	19	0.3	214	27	AJ281464	AJ281464 4A3A-P4F2
520	19	0.3	215	92	AW601182	AW601182 RC1-BT025
c 521	19	0.3	215	160	AQ477810	AQ477810 CITBI-E1-
c 522	19	0.3	216	24	AI714332	AI714332 UI-R-AF1-
c 523	19	0.3	218	98	BB021300	BB021300 BB021300
c 524	19	0.3	219	23	AI630432	AI630432 ad10a11.y
c 525	19	0.3	222	13	AA860897	AA860897 aK22a03.s
526	19	0.3	224	3	AA208476	AA208476 mv85h02.r
527	19	0.3	225	140	C84763	C84763 C84763 Dict
c 528	19	0.3	226	39	AW046850	AW046850 UI-M-BH1-
529	19	0.3	229	14	AA972643	AA972643 oq01b10.s
c 530	19	0.3	229	144	R45698	R45698 yg45h12.s1
531	19	0.3	231	15	AI077164	AI077164 TENU3386
c 532	19	0.3	232	19	AI395122	AI395122 MA002468.

c 533	19	0.3	232	97	BB002109	BB002109 BB002109
c 534	19	0.3	234	128	BB355471	BB355471 BB355471
535	19	0.3	236	35	AV346608	AV346608 AV346608
c 536	19	0.3	238	29	AV017463	AV017463 AV017463
c 537	19	0.3	239	4	AA262753	AA262753 zs24b12.s
538	19	0.3	240	29	AU071592	AU071592 AU071592
539	19	0.3	240	29	AU071844	AU071844 AU071844
c 540	19	0.3	240	29	AU072410	AU072410 AU072410
541	19	0.3	240	29	AU072970	AU072970 AU072970
542	19	0.3	240	94	AW734787	AW734787 ak88g07.y
543	19	0.3	241	138	BF012989	BF012989 ux83b05.y
c 544	19	0.3	242	88	AW270424	AW270424 xp75e03.x
c 545	19	0.3	242	129	BB370055	BB370055 BB370055
c 546	19	0.3	245	101	BB147765	BB147765 BB147765
c 547	19	0.3	245	104	BE144552	BE144552 MR0-HT016
c 548	19	0.3	246	145	T65208	T65208 yc79a04.s1
c 549	19	0.3	247	35	AV323391	AV323391 AV323391
550	19	0.3	248	107	BE345871	BE345871 946026E11
551	19	0.3	249	104	BE107194	BE107194 UI-R-BS1-
552	19	0.3	249	107	BE345739	BE345739 946025E11
c 553	19	0.3	251	15	AI030153	AI030153 UI-R-C0-j
c 554	19	0.3	252	23	AI666443	AI666443 mu16e05.x
c 555	19	0.3	253	36	AV369839	AV369839 AV369839
c 556	19	0.3	254	25	AI839263	AI839263 UI-M-AK0-
c 557	19	0.3	255	32	AV173700	AV173700 AV173700
c 558	19	0.3	256	33	AV271845	AV271845 AV271845
c 559	19	0.3	256	111	BE691146	BE691146 uv56d11.x
c 560	19	0.3	258	89	AW335536	AW335536 S48D12 AG
c 561	19	0.3	258	97	BB002119	BB002119 BB002119
562	19	0.3	259	96	AW904081	AW904081 PM4-NN104
c 563	19	0.3	261	25	AI844738	AI844738 UI-M-AL1-
c 564	19	0.3	261	106	BE335495	BE335495 us03h12.x
565	19	0.3	263	14	AI002732	AI002732 am94b07.s
c 566	19	0.3	266	138	BE979264	BE979264 UI-M-BG2-
c 567	19	0.3	266	144	R49147	R49147 yg69h07.s1
c 568	19	0.3	267	99	BB055020	BB055020 BB055020
c 569	19	0.3	267	131	BB473672	BB473672 BB473672
c 570	19	0.3	267	141	H08236	H08236 y192h05.s1
571	19	0.3	268	111	BE664930	BE664930 152833 MA
c 572	19	0.3	268	183	AZ515432	AZ515432 BMBAC11N0
573	19	0.3	270	89	AW329444	AW329444 N200679e
574	19	0.3	271	130	BB432850	BB432850 BB432850
c 575	19	0.3	271	132	BB496872	BB496872 BB496872
c 576	19	0.3	272	138	BE949175	BE949175 UI-M-BH3-
577	19	0.3	274	29	AU053084	AU053084 AU053084
578	19	0.3	274	37	AV560071	AV560071 AV560071
579	19	0.3	274	105	BE233423	BE233423 139601 MA
580	19	0.3	274	136	BE807272	BE807272 ss15h02.y
c 581	19	0.3	276	104	BE112838	BE112838 UI-R-CA0-
582	19	0.3	277	29	AU062125	AU062125 AU062125
583	19	0.3	277	132	BB527328	BB527328 BB527328

	584	19	0.3	278	93	AW672665 AW672665 ba49b05.y
c	585	19	0.3	278	138	BE956035 BE956035 UI-H-BH1-
	586	19	0.3	280	24	AI707214 AI707214 T7284 MVA
c	587	19	0.3	280	102	BB174836 BB174836 BB174836
c	588	19	0.3	283	19	AI395002 AI395002 MA002009.
c	589	19	0.3	283	26	AI919422 AI919422 tm83f03.x
c	590	19	0.3	285	15	AI029253 AI029253 UI-R-C0-i
	591	19	0.3	285	39	AW075017 AW075017 xa66f12.x
c	592	19	0.3	285	104	BE110117 BE110117 UI-R-BJ1-
c	593	19	0.3	286	21	AI537268 AI537268 tp01h10.x
c	594	19	0.3	286	40	AW138324 AW138324 UI-H-BI1-
	595	19	0.3	286	128	BB358296 BB358296 BB358296
	596	19	0.3	287	30	AV095943 AV095943 AV095943
c	597	19	0.3	288	18	AI305169 AI305169 ql84h08.x
	598	19	0.3	288	24	AI707449 AI707449 T7708 MVA
c	599	19	0.3	289	22	AI579287 AI579287 UI-R-AG0-
c	600	19	0.3	289	96	AW873704 AW873704 hq29a06.x
c	601	19	0.3	291	33	AV258766 AV258766 AV258766
	602	19	0.3	291	128	BB354833 BB354833 BB354833
c	603	19	0.3	292	15	AI030891 AI030891 UI-R-C0-j
	604	19	0.3	293	32	AV204540 AV204540 AV204540
c	605	19	0.3	293	103	BB231466 BB231466 BB231466
	606	19	0.3	294	101	BB137040 BB137040 BB137040
c	607	19	0.3	296	10	AA633465 AA633465 np62d04.s
	608	19	0.3	296	29	AU053738 AU053738 AU053738
c	609	19	0.3	298	21	AI511331 AI511331 UI-R-C3-s
c	610	19	0.3	298	104	BE111554 BE111554 UI-R-BJ1-
	611	19	0.3	299	101	BB137087 BB137087 BB137087
c	612	19	0.3	300	16	AI112715 AI112715 UI-R-Y0-m
	613	19	0.3	300	139	C51443 C51443 C51443 Yuji
c	614	19	0.3	301	20	AI454570 AI454570 UI-R-BT0-
c	615	19	0.3	302	26	AI928215 AI928215 wo95e05.x
c	616	19	0.3	302	103	BB216937 BB216937 BB216937
c	617	19	0.3	302	134	BE097162 BE097162 UI-R-BO1-
	618	19	0.3	303	108	BE439382 EST17 Hum
c	619	19	0.3	306	34	AV304418 AV304418 AV304418
c	620	19	0.3	306	134	BE096361 BE096361 UI-R-BU0-
c	621	19	0.3	307	38	AW016217 AW016217 UI-H-BI0p
	622	19	0.3	308	92	AW577404 AW577404 QV3-BT053
c	623	19	0.3	308	103	BB226251 BB226251 BB226251
c	624	19	0.3	308	172	AQ903062 AQ903062 GSSTc0776
c	625	19	0.3	309	7	AA426011 AA426011 zw49f01.s
	626	19	0.3	309	88	AW281089 AW281089 fj50a05.y
	627	19	0.3	310	10	AA687964 nv13h08.s AA687964
c	628	19	0.3	311	128	BB359124 BB359124 BB359124
	629	19	0.3	312	105	BE192794 BE192794 EST331523
c	630	19	0.3	314	19	AI395106 AI395106 MA002430.
c	631	19	0.3	314	140	F00219 F00219 HSB81G101 S
	632	19	0.3	315	8	AA497030 AA497030 ae32h09.s
	633	19	0.3	315	34	AV307392 AV307392 AV307392
c	634	19	0.3	315	90	AW449728 AW449728 UI-H-BI3-

c	635	19	0.3	316	100	BB106035 BB106035 BB106035
c	636	19	0.3	316	138	BE979161 BE979161 UI-M-BG2-
	637	19	0.3	317	108	BE428808 BE428808 MTD011.A1
c	638	19	0.3	317	138	BE950587 BE950587 UI-M-CE0-
c	639	19	0.3	319	12	AA809560 AA809560 nz17a03.s
c	640	19	0.3	319	25	AI839358 AI839358 UI-M-AN0-
c	641	19	0.3	319	134	BE096996 BE096996 UI-R-BS0-
	642	19	0.3	320	20	AI439728 AI439728 ti55f10.x
c	643	19	0.3	321	16	AI100022 AI100022 34175 Lam
c	644	19	0.3	321	141	H26844 H26844 yl15a04.r1
c	645	19	0.3	323	101	BB138235 BB138235 BB138235
c	646	19	0.3	323	138	BF007783 BF007783 1590000 A
	647	19	0.3	324	132	BB497871 BB497871 BB497871
	648	19	0.3	326	88	AW308742 AW308742 sf71b03.y
c	649	19	0.3	326	104	BE116377 BE116377 UI-R-BS1-
	650	19	0.3	328	20	AI464752 AI464752 mz76b07.y
	651	19	0.3	328	28	AL376278 AL376278 MtBB22E05
c	652	19	0.3	328	39	AW069630 AW069630 cr47h06.x
c	653	19	0.3	328	144	R42442 R42442 yf92h08.s1
c	654	19	0.3	329	94	AW738836 AW738836 gb03a09.y
c	655	19	0.3	329	148	AQ059725 AQ059725 RPC111-49
c	656	19	0.3	332	38	AW016315 AW016315 UI-H-BI0-
	657	19	0.3	334	143	N97467 N97467 0594M3 gmbP
c	658	19	0.3	335	14	AI003432 AI003432 ak87c06.s
c	659	19	0.3	336	8	AA548554 AA548554 nk50b06.s
c	660	19	0.3	336	15	AI083873 AI083873 qf22b09.x
c	661	19	0.3	336	136	BE806980 BE806980 ss09e06.y
c	662	19	0.3	337	25	AI836550 AI836550 UI-M-AP0-
c	663	19	0.3	337	25	AI840946 AI840946 UI-M-AK0-
c	664	19	0.3	338	90	AW451707 AW451707 UI-H-BI3-
c	665	19	0.3	338	90	AW452801 AW452801 UI-H-BI3-
	666	19	0.3	339	111	BE654141 BE654141 UI-M-AN1-
c	667	19	0.3	340	9	AA565169 AA565169 nj36a02.s
c	668	19	0.3	340	87	AW167124 AW167124 xg71c08.x
	669	19	0.3	340	94	AW735440 AW735440 pa54f04.y
c	670	19	0.3	340	146	T98960 T98960 ye66b01.r1
	671	19	0.3	343	39	AW084396 AW084396 xc56e07.x
c	672	19	0.3	343	87	AW232119 AW232119 fj14b08.x
c	673	19	0.3	344	8	AA488844 AA488844 aa55a11.s
	674	19	0.3	347	39	AW021379 AW021379 df21g03.y
c	675	19	0.3	347	138	BE986822 BE986822 UI-M-CG0p
c	676	19	0.3	348	25	AI828393 AI828393 wk84e04.x
c	677	19	0.3	348	137	BE947214 BE947214 UI-M-BH3-
c	678	19	0.3	348	151	AQ278833 CITBI-E1-
c	679	19	0.3	350	23	AI692278 AI692278 wd85a04.x
	680	19	0.3	352	19	AI373359 AI373359 qz45f09.x
c	681	19	0.3	353	16	AI111414 AI111414 UI-R-Y0-m
	682	19	0.3	353	172	AQ903008 AQ903008 GSSTc0789
c	683	19	0.3	354	15	AI065829 AI065829 ag94c05.x
c	684	19	0.3	354	22	AI574785 AI574785 UI-R-G0-u
c	685	19	0.3	354	25	AI844586 AI844586 UI-M-AJ1-

c	686	19	0.3	356	142	H70311 H70311 yr95g08.r1
	687	19	0.3	357	13	AA865595 AA865595 oh51f06.s
c	688	19	0.3	357	134	BE092836 BE092836 CM0-BT074
c	689	19	0.3	357	138	BE982722 UI-M-CG0p
	690	19	0.3	357	172	AQ905483 AQ905483 GSSTc0728
	691	19	0.3	357	172	AQ905957 AQ905957 GSSTc0106
c	692	19	0.3	358	25	AI840395 AI840395 UI-M-AH0-
c	693	19	0.3	360	27	AI954885 AI954885 wx90f09.x
	694	19	0.3	360	91	AW490392 AW490392 UI-M-BH3-
c	695	19	0.3	362	149	AQ120806 AQ120806 HS_3076_A
c	696	19	0.3	362	174	AZ062579 AZ062579 RPCI-23-4
c	697	19	0.3	363	16	AI097106 AI097106 qa03b11.x
c	698	19	0.3	363	22	AI582224 AI582224 tq65a01.x
c	699	19	0.3	365	9	AA582284 AA582284 nn49b06.s
c	700	19	0.3	365	22	AI579285 AI579285 UI-R-AG0-
c	701	19	0.3	365	89	AW323039 AW323039 uo56e04.x
	702	19	0.3	366	15	AI029456 AI029456 UI-R-C0-i
c	703	19	0.3	366	27	AI939354 AI939354 qf22b09.x
	704	19	0.3	366	29	AU069915 AU069915 AU069915
c	705	19	0.3	366	144	R48942 R48942 yg57c05.s1
c	706	19	0.3	367	149	AQ116622 AQ116622 HS_3076_A
	707	19	0.3	367	150	AQ210599 AQ210599 HS_2212_A
c	708	19	0.3	368	15	AI071129 AI071129 UI-R-C2-m
c	709	19	0.3	368	38	AW016684 AW016684 UI-H-BI0p
c	710	19	0.3	369	12	AA858550 AA858550 UI-R-E0-b
	711	19	0.3	369	90	AW415978 AW415978 50611 MAR
	712	19	0.3	369	177	AZ227377 AZ227377 RPCI-23-5
c	713	19	0.3	371	25	AI839305 AI839305 UI-M-AO0-
	714	19	0.3	372	107	BE411135 BE411135 ISC002.D0
c	715	19	0.3	373	143	R02206 R02206 ye83b05.r1
c	716	19	0.3	374	139	C22581 C22981 C22981 Dict
c	717	19	0.3	375	10	AA683306 AA683306 ae90b05.s
c	718	19	0.3	376	11	AA731296 AA731296 nw67h12.s
c	719	19	0.3	376	15	AI052175 AI052175 cy93d03.x
c	720	19	0.3	376	91	AW507243 AW507243 EST00671
	721	19	0.3	376	147	W91141 W91141 mf71c03.r1
c	722	19	0.3	377	142	N35030 N35030 yy17c02.s1
c	723	19	0.3	378	15	AI017831 AI017831 ov08e04.x
	724	19	0.3	381	96	AW912992 AW912992 uf47f04.y
c	725	19	0.3	383	24	AI764362 AI764362 UI-R-Y0-a
c	726	19	0.3	384	8	AA490727 AA490727 aa47b07.r
c	727	19	0.3	384	19	AI395084 AI395084 MA002372.
c	728	19	0.3	385	9	AA553737 AA553737 n113h04.s
c	729	19	0.3	385	23	AI659446 AI659446 tu30d09.x
c	730	19	0.3	385	38	AV646271 AV646271 AV646271
	731	19	0.3	385	89	AW365493 AW365493 MR0-HT007
c	732	19	0.3	385	91	AW533203 AW533203 UI-R-BU0-
	733	19	0.3	385	147	W66691 W66691 me26a05.r1
c	734	19	0.3	386	3	AA150720 AA150720 z146f04.s
	735	19	0.3	386	3	AA188722 AA188722 zp78b03.r
c	736	19	0.3	386	107	BE348032 BE348032 sp10e01.y

c	737	19	0.3	387	7	AA460341 AA460341 zx66g03.s
c	738	19	0.3	390	12	AA805580 AA805580 ob44g09.s
	739	19	0.3	390	12	AA810705 AA810705 oa73d09.s
c	740	19	0.3	390	20	AI415862 AI415862 fb36a07.x
c	741	19	0.3	390	148	AQ055711 AQ055711 CIT-HSP-2
c	742	19	0.3	391	91	AW522768 AW522768 UI-R-BO0-
c	743	19	0.3	392	14	AA999065 AA999065 UI-R-C0-i
	744	19	0.3	392	15	AI073663 AI073663 ox98c11.s
c	745	19	0.3	392	183	B23292 B23292 F28A3TF IGF
	746	19	0.3	393	151	AQ323596 AQ323596 RPCI11-93
c	747	19	0.3	394	105	BE232607 BE232607 137675 MA
	748	19	0.3	395	10	AA646866 AA646866 ve41h01.r
c	749	19	0.3	395	25	AI834925 AI834925 UI-M-AN1-
c	750	19	0.3	395	91	AW489356 AW489356 UI-M-BH3-
c	751	19	0.3	396	142	H67974 H67974 yr76a03.r1
	752	19	0.3	398	8	AA507855 AA507855 ng90b09.s
c	753	19	0.3	398	13	AA891288 AA891288 EST195091
c	754	19	0.3	398	23	AI658604 AI658604 tu21f10.x
c	755	19	0.3	399	13	AA861913 AA861913 ak44f10.s
c	756	19	0.3	399	14	AA956761 AA956761 UI-R-E1-f
c	757	19	0.3	401	143	N97942 N97942 1580C3 czap
c	758	19	0.3	401	181	AZ402980 AZ402980 1M0170K02
c	759	19	0.3	402	13	AA924813 AA924813 UI-R-A1-e
	760	19	0.3	402	21	AI551248 AI551248 vo68a04.x
c	761	19	0.3	402	26	AI928730 AI928730 wp59h07.x
c	762	19	0.3	403	39	AW087848 AW087848 xb38h03.x
	763	19	0.3	403	138	BE982665 BE982665 UI-M-CG0p
c	764	19	0.3	404	12	AA825832 AA825832 od59c03.s
c	765	19	0.3	404	20	AI452759 AI452759 tj45b12.x
c	766	19	0.3	404	40	AW139316 AW139316 UI-H-BI1-
c	767	19	0.3	405	26	AI916618 AI916618 wa28g06.x
c	768	19	0.3	405	40	AW139985 AW139985 UI-H-BI1-
	769	19	0.3	405	111	BE667528 BE667528 151949 MA
	770	19	0.3	405	166	AQ737400 AQ737400 HS_3195_A
c	771	19	0.3	406	141	H58299 H58299 yr20g05.r1
c	772	19	0.3	407	10	AA632745 AA632745 np83f10.s
	773	19	0.3	407	111	BE668236 BE668236 156832 MA
	774	19	0.3	407	193	FR0034416 AL030783 Fugu rubr
c	775	19	0.3	408	90	AW410253 AW410253 fh04f08.y
c	776	19	0.3	408	94	AW699034 AW699034 gb19a07.y
	777	19	0.3	408	146	W17305 W17305 zb17g08.r1
c	778	19	0.3	408	160	AQ495998 AQ495998 HS_5064_B
c	779	19	0.3	409	9	AA602484 AA602484 no89c01.s
	780	19	0.3	409	36	AV409984 AV409984 AV409984
c	781	19	0.3	409	87	AW169387 AW169387 xj25h02.x
c	782	19	0.3	409	138	BE979785 BE979785 UI-M-BG2-
	783	19	0.3	410	149	AQ147301 AQ147301 HS_3054_B
c	784	19	0.3	411	137	BE946638 BE946638 UI-M-BZ0-
c	785	19	0.3	411	144	R60530 R60530 yh13g10.s1
c	786	19	0.3	411	191	CNS02CU8 AL191609 Tetraodon
c	787	19	0.3	412	10	AA687403 AA687403 ns16a03.s

c 788	19	0.3	412	21	AI545458	AI545458 fb81e10.x
c 789	19	0.3	412	134	BE103999	BE103999 UI-R-BX0-
790	19	0.3	413	11	AA777583	AA777583 zi98c08.s
791	19	0.3	414	7	AA447051	AA447051 zw86b05.r
c 792	19	0.3	414	25	AI843364	AI843364 UI-M-AQ1-
c 793	19	0.3	415	39	AW061181	AW061181 UI-M-BH1-
c 794	19	0.3	415	91	AW526961	AW526961 UI-R-BO1-
795	19	0.3	417	19	AI390671	AI390671 mu92b06.y
c 796	19	0.3	417	23	AI667176	AI667176 fc44b08.x
c 797	19	0.3	417	23	AI706107	AI706107 UI-R-AC0-
798	19	0.3	417	140	C92992	C92992 C92992 Dict
799	19	0.3	418	22	AI592354	AI592354 vs72c09.y
800	19	0.3	418	91	AW503190	AW503190 UI-HF-BN0
c 801	19	0.3	418	140	C75048	C75048 C75048 Huma
c 802	19	0.3	418	148	AQ005204	AQ005204 CIT-HSP-2
803	19	0.3	418	150	AQ200312	AQ200312 RPCI11-58
804	19	0.3	420	1	AA017099	AA017099 ze33a10.s
c 805	19	0.3	420	8	AA534078	AA534078 nj97d12.s
c 806	19	0.3	420	19	AI395070	AI395070 MA002334.
c 807	19	0.3	420	156	AQ338403	AQ338403 HS_5011_B
808	19	0.3	421	1	AA049886	AA049886 mj11b06.r
c 809	19	0.3	421	13	AA877518	AA877518 nr01f02.s
810	19	0.3	421	150	AQ216412	AQ216412 HS_2189_A
c 811	19	0.3	421	171	AQ832428	AQ832428 HS_3097_B
c 812	19	0.3	422	3	AA149897	AA149897 zo68e12.s
c 813	19	0.3	422	13	AA875217	AA875217 UI-R-EO-c
c 814	19	0.3	422	90	AW445096	AW445096 UI-H-BI3-
815	19	0.3	423	7	AA424395	AA424395 zv90h07.r
c 816	19	0.3	423	27	AI975144	AI975144 EST269738
817	19	0.3	423	111	BE663567	BE663567 147439 MA
c 818	19	0.3	423	181	AZ400106	AZ400106 1M0166E13
c 819	19	0.3	424	3	AA161093	AA161093 zo58a07.s
820	19	0.3	424	8	AA499710	AA499710 vg05a08.r
c 821	19	0.3	424	11	AA740892	AA740892 nz02e07.s
c 822	19	0.3	424	88	AW274949	AW274949 xm63b09.x
823	19	0.3	424	133	BE022890	BE022890 sm77g08.y
c 824	19	0.3	424	134	BE099374	BE099374 UI-R-BJ1-
c 825	19	0.3	424	138	BE955682	BE955682 UI-M-BH4-
c 826	19	0.3	425	8	AA483830	AA483830 ne41f03.s
827	19	0.3	425	26	AI881000	AI881000 T7768 MVA
c 828	19	0.3	425	142	N35385	N35385 yy23d07.s1
829	19	0.3	425	148	AQ041347	AQ041347 CIT-HSP-2
830	19	0.3	425	148	AQ044353	AQ044353 CIT-HSP-2
c 831	19	0.3	425	192	CN504KVL	AL295338 Tetraodon
832	19	0.3	426	7	AA397638	AA397638 zt79e11.r
c 833	19	0.3	426	7	AA424527	AA424527 zv90h07.s
c 834	19	0.3	427	91	AW529231	AW529231 UI-R-BT1-
835	19	0.3	428	11	AA713857	AA713857 nw04f07.s
c 836	19	0.3	428	110	BE645280	BE645280 7e67e02.x
837	19	0.3	428	193	PR0002851	Z86634 F.rubripes
c 838	19	0.3	429	11	AA741293	AA741293 nz01d09.s

839	19	0.3	429	40	AW153016	AW153016 se33h06.y
c 840	19	0.3	429	141	H11911	H11911 ym10d09.s1
841	19	0.3	430	3	AA153170	AA153170 mr94d11.r
c 842	19	0.3	430	26	AI864491	AI864491 wj93b02.x
c 843	19	0.3	430	134	BE099804	BE099804 UI-R-BJ1-
c 844	19	0.3	431	16	AI090008	AI090008 qb13h03.x
c 845	19	0.3	431	19	AI395426	AI395426 MA003617.
c 846	19	0.3	431	27	AI962764	AI962764 wq43c05.x
847	19	0.3	431	147	W88760	W88760 zh74e05.s1
c 848	19	0.3	433	3	AA191134	AA191134 zp86c08.r
c 849	19	0.3	433	23	AI660376	AI660376 we62d06.x
850	19	0.3	433	147	W74364	W74364 zd60g12.r1
c 851	19	0.3	433	156	AQ359641	AQ359641 HS_5034_A
c 852	19	0.3	434	4	AA245715	AA245715 mx05d11.r
853	19	0.3	434	5	AA306274	AA306274 EST177241
c 854	19	0.3	434	11	AA741032	AA741032 ny99c11.s
c 855	19	0.3	434	39	AW049052	AW049052 UI-M-BH1-
c 856	19	0.3	434	39	AW086050	AW086050 xc76h02.x
c 857	19	0.3	434	138	BE955367	BE955367 UI-M-BH4-
c 858	19	0.3	434	161	AQ569041	AQ569041 HS_5344_A
c 859	19	0.3	435	2	AA141629	AA141629 CK02453.3
c 860	19	0.3	435	88	AW296370	AW296370 UI-H-BW0-
c 861	19	0.3	436	12	AA859355	AA859355 UI-R-EO-c
c 862	19	0.3	436	27	AI970952	AI970952 wr22a04.x
c 863	19	0.3	437	16	AI115998	AI115998 ue25f01.x
c 864	19	0.3	437	18	AI274698	AI274698 qw02d05.x
865	19	0.3	437	20	AI432148	AI432148 tg77b12.x
866	19	0.3	438	24	AI732184	AI732184 ae32h09.x
c 867	19	0.3	439	3	AA159142	AA159142 zo77a08.s
c 868	19	0.3	439	106	BE279899	BE279899 601157275
c 869	19	0.3	440	27	AI952382	AI952382 wx73a07.x
c 870	19	0.3	440	104	BE113555	BE113555 UI-R-CA0-
c 871	19	0.3	441	40	AW124825	AW124825 UI-M-BH2.
c 872	19	0.3	441	134	BE100748	BE100748 UI-R-BJ1-
c 873	19	0.3	441	138	BE979653	BE979653 UI-M-BG2-
874	19	0.3	441	143	N64216	N64216 yv63a07.r1
c 875	19	0.3	441	148	AQ040431	AQ040431 CIT-HSP-2
c 876	19	0.3	442	1	AA055573	AA055573 z175c10.s
c 877	19	0.3	442	13	AA899354	AA899354 UI-R-EO-c
c 878	19	0.3	442	16	AI112372	AI112372 UI-R-Y0-m
c 879	19	0.3	443	89	AW341101	AW341101 xz94c07.x
c 880	19	0.3	443	138	BE980391	BE980391 UI-M-BG2-
881	19	0.3	444	3	AA198216	AA198216 mv44a07.r
882	19	0.3	444	13	AA925420	AA925420 UI-R-A1-e
c 883	19	0.3	445	20	AI428448	AI428448 mo20b03.x
884	19	0.3	445	87	AW231098	AW231098 uo39e07.y
c 885	19	0.3	445	91	AW480248	AW480248 UI-M-BH3-
c 886	19	0.3	445	162	AQ664493	AQ664493 HS_5212_A
c 887	19	0.3	446	8	AA535391	AA535391 nj80e09.s
c 888	19	0.3	446	91	AW480344	AW480344 30919 MAR
c 889	19	0.3	446	91	AW490894	AW490894 UI-M-BH3-

c 890	19	0.3	446	134	BE103729	BE103729 UI-R-BX0-
c 891	19	0.3	446	140	C82678	C82678 C82678 rabb
892	19	0.3	446	140	C83534	C83534 C83534 rabb
893	19	0.3	446	140	C89301	C89301 C89301 Mous
c 894	19	0.3	447	16	AI092696	AI092696 oz89f04.x
c 895	19	0.3	447	142	N24434	N24434 yx26b10.r1
896	19	0.3	448	91	AW535699	AW535699 UI-R-BS0-
897	19	0.3	449	17	AI231597	AI231597 EST228285
c 898	19	0.3	449	25	AI839518	AI839518 UI-M-AN0-
c 899	19	0.3	449	90	AW450860	AW450860 UI-H-BI3-
900	19	0.3	449	97	AW988368	AW988368 ug07f11.y
901	19	0.3	449	108	BE454770	BE454770 HVSMEh009
c 902	19	0.3	450	104	BE116820	BE116820 UI-R-BS1-
c 903	19	0.3	450	141	H07111	H07111 yl82c02.s1
c 904	19	0.3	451	3	AA203980	AA203980 mu28h08.r
c 905	19	0.3	451	109	BE507792	BE507792 dc22g09.y
c 906	19	0.3	453	90	AW410175	AW410175 fh04f04.y
c 907	19	0.3	453	97	AW960553	AW960553 EST372624
c 908	19	0.3	453	138	BE979860	BE979860 UI-M-BG2-
c 909	19	0.3	453	138	BE994968	BE994968 UI-M-CG0p
c 910	19	0.3	454	88	AW298748	AW298748 UI-H-BW0-
c 911	19	0.3	454	91	AW507026	AW507026 707106B12
912	19	0.3	454	160	AQ472456	AQ472456 CITBI-E1-
c 913	19	0.3	455	8	AA504130	AA504130 aa59c07.s
c 914	19	0.3	455	25	AI851940	AI851940 UI-M-BH0-
c 915	19	0.3	456	17	AI182496	AI182496 ue21e12.x
916	19	0.3	456	134	BE031341	BE031341 129876 MA
c 917	19	0.3	457	27	AI963361	AI963361 wt62b12.x
918	19	0.3	457	148	AQ054819	AQ054819 CIT-HSP-2
919	19	0.3	457	161	AQ591368	AQ591368 HS_5410_B
c 920	19	0.3	460	3	AA207148	AA207148 zq84a03.s
921	19	0.3	460	36	AV422604	AV422604 AV422604
c 922	19	0.3	461	19	AI339162	AI339162 qt14h02.x
923	19	0.3	462	4	AA239939	AA239939 mx81h12.r
c 924	19	0.3	462	25	AI818586	AI818586 wk74c04.x
c 925	19	0.3	462	91	AW535032	AW535032 UI-R-BS0-
c 926	19	0.3	463	91	AW535775	AW535775 UI-R-BS0-
927	19	0.3	463	105	BE231564	BE231564 136336 MA
928	19	0.3	463	136	BE848528	BE848528 uw39e03.y
c 929	19	0.3	464	12	AA807167	AA807167 oc36f06.s
c 930	19	0.3	464	15	AI065921	AI065921 aj44b10.x
c 931	19	0.3	464	15	AI072893	AI072893 UI-R-Y0-m
932	19	0.3	464	90	AW406517	AW406517 UI-HF-BL0
c 933	19	0.3	464	91	AW529133	AW529133 UI-R-BO1-
c 934	19	0.3	467	13	AA923852	AA923852 UI-R-A1-d
c 935	19	0.3	467	19	AI395036	AI395036 MA002259.
936	19	0.3	467	29	AU031881	AU031881 AU031881
c 937	19	0.3	467	37	AV561325	AV561325 AV561325
c 938	19	0.3	467	88	AW295274	AW295274 UI-H-BI2-
c 939	19	0.3	467	138	BE994124	BE994124 UI-M-CG0p
940	19	0.3	468	5	AA316936	AA316936 EST188590

c 941	19	0.3	468	151	AQ253217	AQ253217 HS_2008_B
942	19	0.3	469	14	AA982991	AA982991 ub59a10.r
943	19	0.3	469	27	AI945342	AI945342 ba11f11.y
944	19	0.3	469	36	AV405414	AV405414 AV405414
c 945	19	0.3	469	94	AW699804	AW699804 gb32f07.y
c 946	19	0.3	469	136	BE803235	BE803235 ar53f06.y
947	19	0.3	469	149	AQ138746	AQ138746 HS_3071_B
c 948	19	0.3	470	18	AI302638	AI302638 qn48a01.x
c 949	19	0.3	470	89	AW310631	AW310631 sg22d10.x
950	19	0.3	470	94	AW742248	AW742248 up54d08.y
951	19	0.3	470	138	BE948518	BE948518 UI-M-BH3-
c 952	19	0.3	470	173	AQ988012	AQ988012 RPCI-23-3
c 953	19	0.3	471	9	AA602344	AA602344 no89h04.s
954	19	0.3	471	24	AI750702	AI750702 cn05a03.x
955	19	0.3	471	165	AQ775056	AQ775056 HS_3155_A
956	19	0.3	472	27	AJ012879	AJ012879 EHEST009
c 957	19	0.3	472	88	AW282034	AW282034 fj61c07.x
c 958	19	0.3	472	91	AW488058	AW488058 UI-M-BH3-
959	19	0.3	472	97	AW989326	AW989326 ug13g08.y
960	19	0.3	472	150	AQ242922	AQ242922 HS_2061_A
c 961	19	0.3	473	21	AI479224	AI479224 tm55h09.x
c 962	19	0.3	473	25	AI835023	AI835023 UI-M-AI1-
963	19	0.3	473	156	AQ349516	AQ349516 RPCI11-11
c 964	19	0.3	474	12	AA808213	AA808213 oc40e06.s
c 965	19	0.3	474	110	BE629545	BE629545 uu31d11.x
c 966	19	0.3	474	138	BE950984	BE950984 UI-M-BH4-
967	19	0.3	474	141	H54295	H54295 yq93f12.s1
c 968	19	0.3	475	25	AI851046	AI851046 UI-M-BH0-
c 969	19	0.3	475	93	AW628155	AW628155 hh86c11.x
970	19	0.3	475	138	BE991905	BE991905 UI-M-BZ1-
c 971	19	0.3	475	175	AZ089786	AZ089786 RPCI-23-2
c 972	19	0.3	476	39	AW046255	AW046255 UI-M-BH1-
973	19	0.3	476	147	W82548	W82548 mf05e07.r1
c 974	19	0.3	477	150	AQ182412	AQ182412 HS_3058_B
c 975	19	0.3	478	11	AA717989	AA717989 vt99h04.r
c 976	19	0.3	478	16	AI113125	AI113125 UI-R-C2p-
c 977	19	0.3	478	92	AW571350	AW571350 ra19c11.y
c 978	19	0.3	478	97	AW944479	AW944479 SD07027.3
979	19	0.3	478	134	BE098228	BE098228 UI-R-BJ1-
c 980	19	0.3	478	149	AQ134291	AQ134291 HS_3047_B
c 981	19	0.3	479	40	AW119188	AW119188 xd79g11.x
c 982	19	0.3	479	90	AW456604	AW456604 UI-M-BH3-
983	19	0.3	480	36	AV402374	AV402374 AV402374
984	19	0.3	480	140	C96667	C96667 C96667 Rice
c 985	19	0.3	481	24	AI777687	EST258482
986	19	0.3	481	88	AW257697	AW257697 SWOv3MCAM
c 987	19	0.3	481	138	BE988439	BE988439 UI-M-CG0p
988	19	0.3	481	140	C90823	C90823 C90823 Dict
989	19	0.3	482	36	AV401823	AV401823 AV401823
990	19	0.3	482	166	AQ692372	AQ692372 HS_5398_B
c 991	19	0.3	482	181	AZ418914	AZ418914 1M0195N09

```
  c 992      19    0.3    483  2  AA099556                    AA099556 zk86h08.s
  c 993      19    0.3    483  146  W05278                     W05278 za83d05.r1
  c 994      19    0.3    484  88  AW297727                   AW297727 UI-H-BW0-
    995      19    0.3    484  96  AW912318                   AW912318 ur92f12.y
    996      19    0.3    484  147  W48693                     W48693 zc41b10.r1
    997      19    0.3    484  147  W49612                     W49612 zc41c01.r1
    998      19    0.3    485  136  BE863148                   BE863148 UI-M-BH0-
    999      19    0.3    486  1  AA010624                   AA010624 zi09h09.s
 c1000       19    0.3    487  1  AA031434                   AA031434 zk13d10.s
```

ALIGNMENTS

```
RESULT   1
BE577445/c
LOCUS       BE577445      621 bp    mRNA            EST       15-AUG-2000
DEFINITION  L48-2287T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-2287
            5', mRNA sequence.
ACCESSION   BE577445
VERSION     BE577445.1  GI:9827244
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry and Molecular Biology
            Oklahoma State University
            350 Noble Research Center, Stillwater, OK 74078-3035, USA
            Tel: 405-744-6207
            Fax: 405-744-7799
            Email: jcushman@biochem.okstate.edu
            PCR PRimers
            FORWARD: T7
            BACKWARD: T3
            Plate: L48-23  row: H  column: 3
            Seq primer: T3
            High quality sequence stop: 350
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
```



```
                     /clone="L48-2287"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression library
                     , 48 hours NaCl treatment"
                     /tissue_type="Leaf, 48 h 0.4M NaCl"
                     /dev_stage="Six week old"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT      176 a     138 c     141 g     166 t
ORIGIN


  Query Match              0.4%;  Score 23;  DB 110;  Length 621;
  Best Local Similarity   100.0%;  Pred. No. 7.6;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    380 tgattaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||||
Db     33 TGATTAATGAAAAAAAAAAAAAA 11


RESULT   2
BE577469/c
LOCUS       BE577469      646 bp    mRNA            EST       15-AUG-2000
DEFINITION  L48-2315T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-2315
            5', mRNA sequence.
ACCESSION   BE577469
VERSION     BE577469.1  GI:9827268
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry and Molecular Biology
            Oklahoma State University
            350 Noble Research Center, Stillwater, OK 74078-3035, USA
            Tel: 405-744-6207
            Fax: 405-744-7799
            Email: jcushman@biochem.okstate.edu
            PCR PRimers
            FORWARD: T7
            BACKWARD: T3
            Plate: L48-24  row: B  column: 3
            Seq primer: T3
```



```
            High quality sequence stop: 350
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone="L48-2315"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression library
                     , 48 hours NaCl treatment"
                     /tissue_type="Leaf, 48 h 0.4M NaCl"
                     /dev_stage="Six week old"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT      175 a     119 c     137 g     215 t
ORIGIN


  Query Match              0.4%;  Score 23;  DB 110;  Length 646;
  Best Local Similarity   100.0%;  Pred. No. 7.6;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    380 tgattaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||||
Db     25 TGATTAATGAAAAAAAAAAAAAA 3


RESULT   3
BB464908/c
LOCUS       BB464908      260 bp    mRNA            EST       22-JUL-2000
DEFINITION  BB464908 RIKEN full-length enriched, 12 days embryo spinal ganglion
            Mus musculus cDNA clone D130095H16 3', mRNA sequence.
ACCESSION   BB464908
VERSION     BB464908.1  GI:9382097
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
```



```
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .260
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D130095H16"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     spinal ganglion"
                     /tissue_type="spinal ganglion"
                     /dev_stage="12 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
```

pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 66 a 62 c 45 g 87 t
ORIGIN

Query Match 0.4%; Score 22; DB 131; Length 260;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2582 tgtctgataaacttgaaaatgt 2603
||||||||||||||||||||||
Db 180 TGTCTGATAAACTTGAAAATGT 159

RESULT 4
BB485782/c
LOCUS BB485782 277 bp mRNA EST 23-JUL-2000
DEFINITION BB485782 RIKEN full-length enriched, 13 days embryo lung Mus musculus cDNA clone D430033H08 3', mRNA sequence.
ACCESSION BB485782
VERSION BB485782.1 GI:9403391
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 277)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc Natl Acad Sci U S A 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1..277
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D430033H08"
/clone_lib="RIKEN full-length enriched, 13 days embryo lung"
/tissue_type="lung"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 65 a 53 c 39 g 120 t
ORIGIN

Query Match 0.4%; Score 22; DB 131; Length 277;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 381 gattaatgaaaaaaaaaaaaa 402
||||||||||||||||||||||
Db 62 GATTAATGAAAAAAAAAAAAA 41

RESULT 5
BB495792/c
LOCUS BB495792 294 bp mRNA EST 26-JUL-2000
DEFINITION BB495792 RIKEN full-length enriched, 13 days embryo stomach Mus musculus cDNA clone D530050B02 3', mRNA sequence.
ACCESSION BB495792
VERSION BB495792.1 GI:9467975
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 294)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc Natl Acad Sci U S A 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1..294
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D530050B02"
/clone_lib="RIKEN full-length enriched, 13 days embryo stomach"
/tissue_type="stomach"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 85 a 69 c 45 g 95 t
ORIGIN

Query Match 0.4%; Score 22; DB 132; Length 294;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2582 tgtctgataaacttgaaaatgt 2603
||||||||||||||||||||||
Db 215 TGTCTGATAAACTTGAAAATGT 194

RESULT 6
AI593494
LOCUS AI593494 411 bp mRNA EST 21-APR-1999
DEFINITION vu51e02.x1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1194938 3' similar to gb:J04513 HEPARIN-BINDING GROWTH FACTOR PRECURSOR 2 (HUMAN);, mRNA sequence.
ACCESSION AI593494
VERSION AI593494.1 GI:4602542
KEYWORDS EST.
SOURCE house mouse.

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:642034
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 410.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1194938"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      121 a      66 c      89 g     134 t       1 others
ORIGIN

  Query Match             0.4%;  Score 22;  DB 22;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2582 tgtctgataaacttgaaaatgt 2603
           |||||||||||||||||||||||
Db     88 TGTCTGATAAACTTGAAAATGT 109
```



```
RESULT   7
BE237300
LOCUS       BE237300      416 bp    mRNA            EST       10-JUL-2000
DEFINITION  146442 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE237300
VERSION     BE237300.1  GI:9021955
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 46  row: K  column: 15
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 4BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
BASE COUNT      142 a      78 c      69 g     127 t
ORIGIN

  Query Match             0.4%;  Score 22;  DB 105;  Length 416;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy    388 gaaaaaaaaaaaaaatgcttga 409
          |||||||||||||||||||||
Db    243 GAAAAAAAAAAAAAATGCTTGA 264


RESULT   8
BE852924/c
LOCUS       BE852924      483 bp    mRNA            EST       26-SEP-2000
DEFINITION  uw36b10.x1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3418747 3', mRNA sequence.
ACCESSION   BE852924
VERSION     BE852924.1  GI:10311263
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: uw36b10.y1
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1094559
            High quality sequence stop: 297.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3418747"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
```



```
                     Soares and M.Fatima Bonaldo."
BASE COUNT      157 a      84 c     109 g     133 t
ORIGIN

  Query Match             0.4%;  Score 22;  DB 136;  Length 483;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1863 tattgctttttgtttctaaaaaa 1884
           ||||||||||||||||||||||
Db     26 TATTGCTTTTGTTTCTAAAAAA 5


RESULT   9
AW637516/c
LOCUS       AW637516      519 bp    mRNA            EST       03-APR-2000
DEFINITION  bl59d06.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone PBX0059D06 5', mRNA sequence.
ACCESSION   AW637516
VERSION     AW637516.1  GI:7394633
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
            Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
            ,J.W., Bonaldo,M.F. and Soares,M.B.
  TITLE     The NIEHS Xenopus Maternal EST Project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Perry J. Blackshear
            Office of Clinical Research and Laboratory of Signal Transduction
            National Institute of Environmental Health Sciences
            A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
            USA
            Tel: 919 541-4899
            Fax: 919 541-4571
            Email: black009@niehs.nih.gov
            Clone is available through Research Genetics, Inc., 2130 Memorial
            Parkway, Huntsville, AL 35901
             phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
            cdna@resgen.com
             DNA Sequencing and analyses performed by National Institutes of
            Health Intramural Sequencing Center (NISC).
            PCR PRimers
            FORWARD: TGTAAAACGACGGCCAGT
            BACKWARD: CAGGAAACAGCTATGACC
            Plate: 0059  row: D  column: 06
```

```
            Seq primer: T7 primer.
FEATURES              Location/Qualifiers
     source           1. .519
                      /organism="Xenopus laevis"
                      /db_xref="taxon:8355"
                      /clone="PBX0059D06"
                      /clone_lib="Blackshear/Soares normalized Xenopus egg
                      library"
                      /sex="female"
                      /tissue_type="unfertilized egg"
                      /cell_type="unfertilized egg"
                      /dev_stage="unfertilized egg"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                      PolyA-selected mRNA was prepared from unfertilized Xenopus
                      laevis eggs.  The library was constructed in the vector
                      pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
                      Soares, M.B. 'Normalization and subtraction:  two
                      approaches to facilitate gene discovery',  Genome Research
                      6:791-806, 1996.  The first strand synthesis used a
                      NotI-dT18 primer; double stranded cDNAs were ligated to
                      EcoRI adapters, digested with NotI, and directionally
                      cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
                      The library contained approximately 7.2 X 10^5
                      recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT       133 a     102 c     109 g     175 t
ORIGIN


  Query Match             0.4%;  Score 22;  DB 93;  Length 519;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaatgct 406
          ||||||||||||||||||||||
Db    247 AATGAAAAAAAAAAAAAATGCT 226


RESULT  10
AV385015
LOCUS       AV385015       539 bp     mRNA            EST       27-OCT-1999
DEFINITION  AV385015 Halocynthia roretzi Fertilized egg Halocynthia roretzi
            cDNA clone 004C08_3 3', mRNA sequence.
ACCESSION   AV385015
VERSION     AV385015.1  GI:6130072
KEYWORDS    EST.
SOURCE      Halocynthia roretzi.
  ORGANISM  Halocynthia roretzi
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Stolidobranchia; Pyuridae; Halocynthia.
REFERENCE   1  (bases 1 to 539)
```



```
  AUTHORS   Makabe,K.W.
  TITLE     Halocynthia roretzi EST
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Kazuhiro W. Makabe
            Department of Zoology, Graduate School of Science
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4095
            Fax: 81-75-705-1113
            Email: kwmakabe@ascidian.zool.kyoto-u.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .539
                      /organism="Halocynthia roretzi"
                      /db_xref="taxon:7729"
                      /clone="004C08_3"
                      /clone_lib="Halocynthia roretzi Fertilized egg"
                      /dev_stage="Fertilized egg"
BASE COUNT       216 a     89 c     116 g     118 t
ORIGIN


  Query Match             0.4%;  Score 22;  DB 36;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    381 gattaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db    514 GATTAATGAAAAAAAAAAAAA 535


RESULT  11
AZ384548/c
LOCUS       AZ384548       592 bp     DNA             GSS       02-OCT-2000
DEFINITION  1M0142C11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0142C11 R, DNA sequence.
ACCESSION   AZ384548
VERSION     AZ384548.1  GI:10498248
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
```



```
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0142  row: C  column: 11
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 592.
FEATURES              Location/Qualifiers
     source           1. .592
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0142C11"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT       161 a     141 c     113 g     177 t
ORIGIN


  Query Match             0.4%;  Score 22;  DB 180;  Length 592;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    951 aaattttatggaaaatatcata 972
          ||||||||||||||||||||||
Db    193 AAATTTTATGGAAAATATCATA 172
```



```
RESULT  12
BE963593/c
LOCUS       BE963593       1026 bp     mRNA            EST       04-OCT-2000
DEFINITION  601657394R1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866278 3',
            mRNA sequence.
ACCESSION   BE963593
VERSION     BE963593.1  GI:10574298
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1026)
  AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM597  row: 1  column: 23
            High quality sequence stop: 31.
FEATURES              Location/Qualifiers
     source           1. .1026
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3866278"
                      /clone_lib="NIH_MGC_67"
                      /tissue_type="retinoblastoma"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.75 kb.   Library constructed by Life
                      Technologies."
BASE COUNT       260 a     289 c     243 g     233 t      1 others
ORIGIN


  Query Match             0.4%;  Score 22;  DB 138;  Length 1026;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    381 gattaatgaaaaaaaaaaaaa 402
```

```
              ||||||||||||||||||||||
Db      42 GATTAATGAAAAAAAAAAAAA 21


RESULT  13
AL386441
LOCUS       AL386441      150 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBC34F09R1 MtBC Medicago truncatula cDNA clone MtBC34F09 T7, mRNA
            sequence.
ACCESSION   AL386441
VERSION     AL386441.1  GI:9686192
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from endomycorrhizal roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBC34F09"
                     /clone_lib="MtBC"
                     /tissue_type="arbuscular mycorrhiza"
                     /dev_stage="harvested 3 weeks post inoculation with Glomus
                     intraradices"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; M. truncatula sterilised seeds were germinated for
                     72h at 25 C, before transplanting into a 1/3 Epoisses soil
                     : 2/3 calcined Terragreen mix in the presence of onion
                     root fragments colonized by the arbuscular mycorrhizal
                     fungus Glomus intraradices (Schenck & Smith, isolate LPA8
                     ). The plants were watered every day and twice a week with
                     a modified nutrient Long Ashton solution without phosphate
                     but with a high level of nitrate. After 3 weeks RNA was
```



```
                     extracted from whole root systems. cDNA was prepared from
                     polyA+ enriched RNA. The cDNA was directionally ligated
                     into Uni-zap XR vector from Stratagene and packaged using
                     Gigapack Gold packaging extracts. Plasmids containing cDNA
                     inserts were mass-excised from phage stocks using ExAssit
                     helper phage and propagated in SOLR cells. Clone ordering
                     and sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France). Note : EST may be of
                     fungal origin."
BASE COUNT       63 a      11 c      25 g      51 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db    114 ATTAATGAAAAAAAAAAAAA 134


RESULT  14
AQ444194/c
LOCUS       AQ444194      157 bp    DNA             GSS       22-SEP-2000
DEFINITION  GSSTc0245 Trypanosoma cruzi random genomic library Trypanosoma
            cruzi genomic clone G1002, DNA sequence.
ACCESSION   AQ444194
VERSION     AQ444194.3  GI:10140125
KEYWORDS    GSS.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 157)
  AUTHORS   Aguero,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.
  TITLE     A random sequencing approach for the analysis of the Trypanosoma
            cruzi genome: general structure, large gene and repetitive DNA
            families and gene discovery
  JOURNAL   Unpublished (2000)
COMMENT     On Sep 14, 2000 this sequence version replaced gi:9372143.
            Contact: Sanchez D.O.
            Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
            San Martin)
            Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
            CP(1650) San Martin, Prov. de BS AS. Argentina
            Tel: 54-11-4580-7255 ext 309
            Fax: 54-11-4752-9639
            Email: dsanchez@iib.unsam.edu.ar
            Sequences were basecalled with phred and vector was masked with
            crossmatch (see http://genome.washington.edu). Sequences were then
```



```
            trimmed from both ends to remove low quality bases and masked
            vector.
            Seq primer: T7
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .157
                     /organism="Trypanosoma cruzi"
                     /strain="CL-Brener"
                     /db_xref="taxon:5693"
                     /clone="G1002"
                     /clone_lib="Trypanosoma cruzi random genomic library"
                     /cell_type="epimastigote"
                     /note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
                     randomly sheared using a nebulizer and the 1 to 2 Kb range
                     was gel purified and cloned into the dephosphoryated
                     HincII site of the vector"
BASE COUNT       41 a      30 c      24 g      62 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 157;  Length 157;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3283 aaaaaggtaaatgcaaaaatt 3303
          |||||||||||||||||||||
Db     59 AAAAAGGTAAATGCAAAAATT 39


RESULT  15
AI054251/c
LOCUS       AI054251      172 bp    mRNA            EST       15-JUL-1998
DEFINITION  qi77b11.x1 NCI_CGAP_Ov26 Homo sapiens cDNA clone IMAGE:1862493 3',
            mRNA sequence.
ACCESSION   AI054251
VERSION     AI054251.1  GI:3322038
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
```



```
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .172
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1862493"
                     /clone_lib="NCI_CGAP_Ov26"
                     /sex="female"
                     /tissue_type="papillary serous carcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP1; mRNA made from
                     papillary serous ovarian carcinoma, cDNA made by oligo-dT
                     priming.  Directionally cloned.  Size-selected on agarose
                     gel, average insert size 500 bp.  Primary library,
                     non-amplified.  "
BASE COUNT       38 a      21 c      25 g      88 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 15;  Length 172;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db     27 ATTAATGAAAAAAAAAAAAA 7


RESULT  16
AU092452
LOCUS       AU092452      176 bp    mRNA            EST       14-JUN-2000
DEFINITION  AU092452 Rice callus Oryza sativa cDNA clone C50919, mRNA sequence.
ACCESSION   AU092452
VERSION     AU092452.1  GI:8527637
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 176)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from callus (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
```

```
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
            C50919_6Z.
FEATURES              Location/Qualifiers
     source           1. .176
                      /organism="Oryza sativa"
                      /strain="cultivar Nipponbare, sub_species Japonica"
                      /db_xref="taxon:4530"
                      /clone="C50919"
                      /clone_lib="Rice callus"
                      /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                      NotI; cDNA prepared from rice callus mRNAs by using
                      oligo(dT) as a primer and ligating to the SalI-NotI site
                      of pBluescript II SK+ phagemid. "
BASE COUNT       64 a     20 c     29 g     55 t      8 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 29;  Length 176;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db    137 ATTAATGAAAAAAAAAAAAA 157


RESULT 17
AI836709/c
LOCUS       AI836709      185 bp    mRNA            EST       14-JUL-1999
DEFINITION  UI-M-AJ0-aax-c-09-0-UI.s1 NIH_BMAP_MOB Mus musculus cDNA clone
            UI-M-AJ0-aax-c-09-0-UI 3', mRNA sequence.
ACCESSION   AI836709
VERSION     AI836709.1  GI:5470922
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 185)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
```



```
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            The following repetitive elements were found in this cDNA sequence:
            29-51, >POLY_A#Simple_repeat
            Seq primer: M13 Forward
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .185
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-AJ0-aax-c-09-0-UI"
                      /clone_lib="NIH_BMAP_MOB"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_MOB library is a non-normalized library
                      constructed from  mouse olfactory bulbs. The tag is a
                      string of 5 nucleotides present  between the Not I site
                      and the oligo-dT track.  The library was  constructed as
                      described by Bonaldo, Lennon and Soares,  Genome Research
                      6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich,
                      Zivic-Miller Laboratories.
                      TAG_LIB=NIH_BMAP_MOB
                      TAG_TISSUE=corpus-striatum
                      TAG_SEQ=ACGGC"
BASE COUNT       40 a     40 c     40 g     65 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 25;  Length 185;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db     58 ATTAATGAAAAAAAAAAAAA 38
```



```
RESULT 18
AW911034
LOCUS       AW911034      213 bp    mRNA            EST       25-MAY-2000
DEFINITION  ur77d09.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3156305 5',
            mRNA sequence.
ACCESSION   AW911034
VERSION     AW911034.1  GI:8076279
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 213)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: ur77d09.x1
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1059061
            Seq primer: -40RP from Gibco
            High quality sequence stop: 197.
FEATURES              Location/Qualifiers
     source           1. .213
                      /organism="Mus musculus"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3156305"
                      /clone_lib="NCI_CGAP_Mam6"
                      /sex="female, virgin"
                      /tissue_type="infiltrating ductal carcinoma"
                      /dev_stage="5 months"
                      /lab_host="DH10B"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigator
                      providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT       84 a     39 c     52 g     37 t      1 others
ORIGIN
```



```
  Query Match             0.4%;  Score 21;  DB 96;  Length 213;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2403 tttagagaaagaagtagagaa 2423
          |||||||||||||||||||||
Db     84 TTTAGAGAAAGAAGTAGAGAA 104


RESULT 19
AI835982/c
LOCUS       AI835982      230 bp    mRNA            EST       14-JUL-1999
DEFINITION  UI-M-AQ0-aah-d-10-0-UI.s2 NIH_BMAP_MHI Mus musculus cDNA clone
            UI-M-AQ0-aah-d-10-0-UI 3', mRNA sequence.
ACCESSION   AI835982
VERSION     AI835982.1  GI:5470235
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 230)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized hippocampus library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .230
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
```

```
                     /db_xref="taxon:10090"
                     /clone="UI-M-AQ0-aah-d-10-0-UI"
                     /clone_lib="NIH_BMAP_MHI"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MHI library is a non-normalized library
                     constructed from  mouse hippocampus. The tag is a string
                     of 5 nucleotides present  between the Not I site and the
                     oligo-dT track.  The library was  constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996. Tissue provided by Ms. Annie Novakovich,
                     Zivic-Miller Laboratories.
                     TAG_LIB=NIH_BMAP_MHI
                     TAG_TISSUE=hippocampus
                     TAG_SEQ=TTCGA"
BASE COUNT       76 a      57 c      36 g      61 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 25;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      23 ATTAATGAAAAAAAAAAAAA 3


RESULT  20
BE751829
LOCUS       BE751829      233 bp    mRNA            EST       15-SEP-2000
DEFINITION  204054 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE751829
VERSION     BE751829.1  GI:10165821
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
```



```
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smithv@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 44  row: H  column: 23
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES                 Location/Qualifiers
     source              1. .233
                         /organism="Bos taurus"
                         /db_xref="taxon:9913"
                         /clone_lib="MARC 2BOV"
                         /tissue_type="pooled"
                         /lab_host="DH10B"
                         /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                         Library made from pooled tissue from testis, thymus,
                         semitendonosus muscle, longissimus muscle, pancreas,
                         adrenal, and endometrium."
BASE COUNT       74 a      42 c      48 g      69 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 135;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db     209 ATTAATGAAAAAAAAAAAAA 229


RESULT  21
AV290224/c
LOCUS       AV290224      237 bp    mRNA            EST       09-NOV-1999
DEFINITION  AV290224 RIKEN full-length enriched, 10 days neonate intestine Mus
            musculus cDNA clone 5133400D01 3' similar to NM_005538 Homo sapiens
            inhibin, beta C (INHBC) mRNA, mRNA sequence.
ACCESSION   AV290224
VERSION     AV290224.1  GI:6304255
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 237)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
```



```
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Misuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN House ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
            "Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            "Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            "High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            "Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES                 Location/Qualifiers
     source              1. .237
                         /organism="Mus musculus"
                         /strain="C57BL/6J"
                         /db_xref="taxon:10090"
                         /clone="5133400D01"
                         /clone_lib="RIKEN full-length enriched, 10 days neonate
                         intestine"
                         /sex="mixed"
                         /tissue_type="intestine"
                         /dev_stage="10 days neonate"
                         /lab_host="DH10B"
                         /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                         prepared and sequenced in Mouse Genome Encyclopedia
                         Project of Genome Exploration Research Group in Riken
```



```
                         Genomic Sciences Center and Genome Science Laboratory in
                         RIKEN. Division of Experimental Animal Research in Riken
                         contributed to prepare mouse tissues. 1st strand cDNA was
                         primed with a primer [5'
                         GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                         prepared by using trehalose thermo-activated reverse
                         transcriptase and subsequently enriched for full-length by
                         cap-trapper. cDNA went through one round of normalization
                         to Rot = 10.0 and subtraction to Rot = 100.0. Second
                         strand cDNA was prepared with the primer adapter of
                         sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                         3']. cDNA was cloned into the XhoI and BamHI sites.
                         Vector: a modified pBluescript KS(+) after bulk excision
                         from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                         BamHI."
BASE COUNT       60 a      32 c      37 g     108 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 34;  Length 237;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db     167 ATTAATGAAAAAAAAAAAAA 147


RESULT  22
CNS00Q5F/c
LOCUS       CNS00Q5F      247 bp    DNA             GSS       28-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence SP6 end of BAC F10K5 of
            IGF library from strain Columbia of Arabidopsis thaliana, genomic
            survey sequence.
ACCESSION   AL085377
VERSION     AL085377.1  GI:5286517
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 247)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
```

```
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES              Location/Qualifiers
     source           1. .247
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /clone_lib="IGF"
                      /clone="P10K5"
                      /note="end : SP6"
BASE COUNT       50 a      52 c      61 g      84 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 190;  Length 247;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1800 aactgagattaaagatttcca 1820
           ||||||||||||||||||||
Db    140 AACTGAGATTAAAGATTTCCA 120


RESULT  23
AA997094/c
LOCUS       AA997094      254 bp    mRNA            EST       04-JUL-1999
DEFINITION  UI-R-C0-hr-c-03-0-UI.s1 UI-R-C0 Rattus norvegicus cDNA clone
            UI-R-C0-hr-c-03-0-UI 3', mRNA sequence.
ACCESSION   AA997094
VERSION     AA997094.1  GI:4281441
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187955.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
```



```
            oligo-dT track served to identify it as a clone from the normalized
            adult Ovary library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics This clone is also available through the I.M.A.G.E.
            Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1781804 The
            following repetitive elements were found in this cDNA sequence:
            60-95, >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .254
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-C0-hr-c-03-0-UI"
                      /clone_lib="UI-R-C0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
                      library is a subtracted library derived from the UI-R-A1
                      and UI-R-E1 libraries. The UI-R-A1 library consisted of a
                      mixture of individually tagged normalized libraries
                      constructed from rat  placenta, adult lung, brain, liver,
                      kidney, heart, spleen, ovary,  and muscle. The UI-R-E1
                      library  consisted of a mixture of individually tagged
                      normalized libraries constructed from 8, 12 and 18-day
                      embryo. The tag is a string of 3-5 nucleotides present
                      between the Not I  site and the oligo-dT track which
                      allows identification of the library of origin of a clone
                      within the mixture. The subtracted library (UI-R-C0) was
                      constructed as follows: PCR amplified cDNA inserts from a
                      pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
                      been derived was used as a driver in a hybridization with
                      the pooled UI-R-A1 and UI-R-E1 library in the form of
                      single-stranded circles.  The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-C0
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
BASE COUNT       81 a      57 c      32 g      84 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 14;  Length 254;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy    382 attaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||||
Db     24 ATTAATGAAAAAAAAAAAAAA 4


RESULT  24
AI707033/c
LOCUS       AI707033      254 bp    mRNA            EST       03-JUN-1999
DEFINITION  UI-R-AD1-zm-e-07-0-UI.s1 UI-R-AD1 Rattus norvegicus cDNA clone
            UI-R-AD1-zm-e-07-0-UI 3', mRNA sequence.
ACCESSION   AI707033
VERSION     AI707033.1  GI:4994933
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized atrium at 15 dpc library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 60-95,
            >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .254
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-AD1-zm-e-07-0-UI"
                      /clone_lib="UI-R-AD1"
                      /dev_stage="adult"
```



```
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AD1
                      library is a normalized library constructed from  15 dpc
                      rat atrium.  The tag is a string of 5 nucleotides present
                      between the Not I site and the oligo-dT track.  The
                      library was  constructed as described by Bonaldo, Lennon
                      and Soares,  Genome Research 6: 791-806, 1996. Tissue
                      provided by Jim Lin, Department of Biology, University of
                      Iowa.
                      TAG_LIB=UI-R-AD1
                      TAG_TISSUE=atrium at 15 dpc
                      TAG_SEQ=GAAGC"
BASE COUNT       77 a      57 c      35 g      84 t       1 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 24;  Length 254;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||||
Db     24 ATTAATGAAAAAAAAAAAAAA 4


RESULT  25
AL047499/c
LOCUS       AL047499      262 bp    mRNA            EST       01-MAR-2000
DEFINITION  DKFZp586A2321_s1 586 (synonym: hutel) Homo sapiens cDNA clone
            DKFZp586A2321, mRNA sequence.
ACCESSION   AL047499
VERSION     AL047499.1  GI:4728495
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 262)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Ottenwaelder, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ottenwaelder B
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 3' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. r1 sequence
```

```
            also available.
            This clone (DKFZp586A2321) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .262
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp586A2321"
                     /clone_lib="586 (synonym: hutel)"
                     /tissue_type="uterus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI/MluI"
BASE COUNT       91 a      32 c      23 g     116 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 262;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      28 ATTAATGAAAAAAAAAAAAA 8


RESULT  26
BB251234/c
LOCUS       BB251234       262 bp    mRNA            EST       06-JUL-2000
DEFINITION  BB251234 RIKEN full-length enriched, 7 days neonate cerebellum Mus
            musculus cDNA clone A730045G06 3', mRNA sequence.
ACCESSION   BB251234
VERSION     BB251234.1  GI:8943980
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 262)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kumakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
```



```
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .262
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A730045G06"
                     /clone_lib="RIKEN full-length enriched, 7 days neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="7 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
```



```
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       85 a      59 c      19 g      99 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 126;  Length 262;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     366 ttcaatattaataatgattaa 386
           |||||||||||||||||||||
Db     123 TTCAATATTAATAATGATTAA 103


RESULT  27
AL047498
LOCUS       AL047498       265 bp    mRNA            EST       01-MAR-2000
DEFINITION  DKFZp586A2321_r1 586 (synonym: hutel),Homo sapiens cDNA clone
            DKFZp586A2321, mRNA sequence.
ACCESSION   AL047498
VERSION     AL047498.3  GI:5936374
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Ottenwaelder, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Apr 30, 1999 this sequence version replaced gi:5866813.
            Contact: Ottenwaelder B
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. sl sequence
            also available.
            This clone (DKFZp586A2321) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .265
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```



```
                     /clone="DKFZp586A2321"
                     /clone_lib="586 (synonym: hutel)"
                     /tissue_type="uterus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI/MluI"
BASE COUNT      120 a      23 c      33 g      89 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 265;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db     235 ATTAATGAAAAAAAAAAAAA 255


RESULT  28
AW522628/c
LOCUS       AW522628       271 bp    mRNA            EST       06-MAR-2000
DEFINITION  UI-R-BO0-ahd-h-12-0-UI.s1 UI-R-BO0 Rattus norvegicus cDNA clone
            UI-R-BO0-ahd-h-12-0-UI 3', mRNA sequence.
ACCESSION   AW522628
VERSION     AW522628.1  GI:7165013
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized thalamus library cDNA Library Preparation: M.B. Soares
```

```
            Lab Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA="Yes.
  FEATURES             Location/Qualifiers
       source          1. .271
                       /organism="Rattus norvegicus"
                       /strain="Sprague-Dawley"
                       /db_xref="taxon:10116"
                       /clone="UI-R-B00-ahd-h-12-0-UI"
                       /clone_lib="UI-R-B00"
                       /dev_stage="adult"
                       /lab_host="DH10B (Life Technologies)"
                       /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                       polylinker; Site_1: Not I; Site_2: Eco RI; The library
                       (UI-R-B00) is a subtracted library derived from a mixture
                       of the following tissues: thalamus, cerebellum,
                       hypothalamus, medulla, pons, midbrain, cerebral cortex,
                       corpus striatum and hippocampus. For a detailed
                       description of the library from which this clone was
                       derived, please visit our web site at
                       ratest.eng.uiowa.edu. The subtraction has been previously
                       described in (Bonaldo, Lennon and Soares, Genome Research
                       6:791-806, 1996)
                       TAG_LIB=UI-R-B00
                       TAG_TISSUE=thalamus
                       TAG_SEQ=GATCG"
  BASE COUNT       59 a      66 c      80 g      66 t
  ORIGIN


    Query Match             0.4%;  Score 21;  DB 91;  Length 271;
    Best Local Similarity   100.0%;  Pred. No. 71;
    Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy     382 attaatgaaaaaaaaaaaaa 402
             |||||||||||||||||||||
  Db      24 ATTAATGAAAAAAAAAAAAA 4


  RESULT  29
  BB133253
  LOCUS       BB133253       289 bp     mRNA            EST       28-JUN-2000
  DEFINITION  BB133253 RIKEN full-length enriched, adult male bone Mus musculus
              cDNA clone 9830006K22 3', mRNA sequence.
  ACCESSION   BB133253
  VERSION     BB133253.1  GI:8787643
  KEYWORDS    EST.
  SOURCE      house mouse.
    ORGANISM  Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



```
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE   1  (bases 1 to 289)
    AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
              ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
              Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
              Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
              Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
              , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
              , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
              ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
              , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
              ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
              , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
              ,M., Muramatsu,M. and Hayashizaki,Y.
    TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
    JOURNAL   Unpublished (2000)
  COMMENT     Contact: Yoshihide Hayashizaki
              Genome Exploration Research Group, Life Science Tsukuba Center,
              Genome Science Laboratory
              The Institute of Physical and Chemical Research (RIKEN), Genomic
              Sciences Center
              3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
              Tel: +81-298-36-9013
              Fax: +81-298-36-9098
              Email: genome-res@rtc.riken.go.jp,
              URL:http://genome.rtc.riken.go.jp/
              Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
              ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
              Thermostabilization and thermoactivation of thermolabile enzymes by
              trehalose and its application for the synthesis of full length
              cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
              Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
              Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
              ,Y. and Hayashizaki,Y.
              Automated filtration-based high-throughput plasmid preparation
              system. Genome Res. 9 (5), 463-470 (1999)
              Carninci,P. and Hayashizaki,Y.
              High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
              19-44 (1999)
              Please visit our web site (http://genome.rtc.riken.go.jp) for
              further details.
  FEATURES             Location/Qualifiers
       source          1. .289
                       /organism="Mus musculus"
                       /db_xref="taxon:10090"
                       /clone="9830006K22"
                       /clone_lib="RIKEN full-length enriched, adult male bone"
                       /sex="male"
                       /tissue_type="bone"
                       /dev_stage="adult"
                       /lab_host="DH10B"
```



```
                       /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                       prepared and sequenced in Mouse Genome Encyclopedia
                       Project of Genome Exploration Research Group in Riken
                       Genomic Sciences Center and Genome Science Laboratory in
                       RIKEN. Division of Experimental Animal Research in Riken
                       contributed to prepare mouse tissues. 1st strand cDNA was
                       primed with a primer (5'
                       GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'), cDNA was
                       prepared by using trehalose thermo-activated reverse
                       transcriptase and subsequently enriched for full-length by
                       cap-trapper. cDNA went through one round of normalization
                       to Rot = 10.0 and subtraction to Rot = 185.0. Second
                       strand cDNA was prepared with the primer adapter of
                       sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                       3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                       modified pBluescript KS(+) after bulk excision from Lambda
                       FLC I."
  BASE COUNT       87 a      57 c      46 g      99 t
  ORIGIN


    Query Match             0.4%;  Score 21;  DB 101;  Length 289;
    Best Local Similarity   100.0%;  Pred. No. 71;
    Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy    2620 gactttgataaatcttttgat 2640
             |||||||||||||||||||||
  Db      79 GACTTTGATAAATCTTTTGAT 99


  RESULT  30
  C97838/c
  LOCUS       C97838         292 bp     mRNA            EST       19-OCT-1998
  DEFINITION  C97838 Rice callus Oryza sativa cDNA clone C63796_1Z, mRNA
              sequence.
  ACCESSION   C97838
  VERSION     C97838.1  GI:3760584
  KEYWORDS    EST.
  SOURCE      Oryza sativa.
    ORGANISM  Oryza sativa
              Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
              Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
  REFERENCE   1  (bases 1 to 292)
    AUTHORS   Sasaki,T. and Minobe,Y.
    TITLE     Rice cDNA from callus
    JOURNAL   Unpublished (1994)
  COMMENT     Contact: Takuji Sasaki
              National Institute of Agrobiological Resources
              Rice Genome Research Program
              2-1-2 Kannondai,Tsukuba
              Ibaraki,
```



```
              Japan 305
              Tel: 0298-38-7441
              Fax: 0298-38-7468
              Email: tsasaki@abr.affrc.go.jp
              PROJECT ='RGP'.
  FEATURES             Location/Qualifiers
       source          1. .292
                       /organism="Oryza sativa"
                       /strain="cultivar Nipponbare, sub_species Japonica"
                       /db_xref="taxon:4530"
                       /clone="C63796_1Z"
                       /clone_lib="Rice callus"
                       /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                       NotI; cDNA prepared from rice callus mRNAs by using
                       oligo(dT) as a primer and ligating to the SalI-NotI site
                       of pBluescript II SK+ phagemid. "
  BASE COUNT       60 a      65 c      71 g      93 t       3 others
  ORIGIN


    Query Match             0.4%;  Score 21;  DB 140;  Length 292;
    Best Local Similarity   100.0%;  Pred. No. 71;
    Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy    1977 agataggggagaaaaatgttac 1997
             |||||||||||||||||||||
  Db     235 AGATAGGGAGAAAAATGTTAC 215


  RESULT  31
  AU108172
  LOCUS       AU108172       314 bp     mRNA            EST       29-SEP-2000
  DEFINITION  AU108172 Rice panicle at flowering stage Oryza sativa cDNA clone
              E3787, mRNA sequence.
  ACCESSION   AU108172
  VERSION     AU108172.1  GI:10423626
  KEYWORDS    EST.
  SOURCE      Oryza sativa.
    ORGANISM  Oryza sativa
              Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
              Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
  REFERENCE   1  (bases 1 to 314)
    AUTHORS   Sasaki,T. and Yamamoto,K.
    TITLE     Rice cDNA from panicle at flowering stage (2000)
    JOURNAL   Unpublished (2000)
  COMMENT     Contact: Takuji Sasaki
              National Institute of Agrobiological Resources
              Rice Genome Research Program
              2-1-2 Kannondai,Tsukuba
              Ibaraki,
              Japan 305
```

```
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT= 'RGP'.
            E3787_1A.
FEATURES             Location/Qualifiers
     source          1. .314
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="E3787"
                     /clone_lib="Rice panicle at flowering stage"
                     /dev_stage="flowering stage"
                     /note="Organ: panicle; Rice cDNA from panicle at flowering
                     stage"
BASE COUNT      102 a     69 c     58 g     83 t      2 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 29;  Length 314;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||||
Db     293 ATTAATGAAAAAAAAAAAAAA 313


RESULT  32
AI111452/c
LOCUS       AI111452      323 bp    mRNA            EST       11-FEB-1999
DEFINITION  UI-R-Y0-lz-c-12-0-UI.a1 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-lz-c-12-0-UI 3', mRNA sequence.
ACCESSION   AI111452
VERSION     AI111452.1  GI:3511494
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
```



```
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult Eye library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics The following repetitive elements were found in this cDNA
            sequence: 60-95, >AT_rich#Low_complexity
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-Y0-lz-c-12-0-UI"
                     /clone_lib="UI-R-Y0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
                     library is a subtracted library derived from an
                     individually-tagged normalized whole-eye (minus the lens)
                     library. The driver for the subtraction consisted of a
                     pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
                     UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
                     3-5 nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-Y0) was constructed as follows: PCR
                     amplified cDNA inserts from previous library clones from
                     which 3' ESTs had  been derived were used as a driver in a
                     hybridization with the normalized  whole-eye library in
                     the form of single-stranded circles. The remaining
                     single-stranded circles (subtracted library) was purified
                     by hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-Y0
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)"
BASE COUNT      104 a     76 c     46 g     97 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 16;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
```



```
Db      24 ATTAATGAAAAAAAAAAAAA 4


RESULT  33
AI764576/c
LOCUS       AI764576      323 bp    mRNA            EST       25-JUN-1999
DEFINITION  UI-R-Y0-abv-e-11-0-UI.s2 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-abv-e-11-0-UI 3', mRNA sequence.
ACCESSION   AI764576
VERSION     AI764576.1  GI:5210511
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized Eye library cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.reagen.com) The following repetitive elements were
            found in this cDNA sequence: 60-95, >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-Y0-abv-e-11-0-UI"
                     /clone_lib="UI-R-Y0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
```



```
                     library is a subtracted library derived from an
                     individually-tagged normalized whole-eye (minus the lens)
                     library. The driver for the subtraction consisted of a
                     pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
                     UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
                     3-5 nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-Y0) was constructed as follows: PCR
                     amplified cDNA inserts from previous library clones from
                     which 3' ESTs had  been derived were used as a driver in a
                     hybridization with the normalized  whole-eye library in
                     the form of single-stranded circles. The remaining
                     single-stranded circles (subtracted library) was purified
                     by hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-Y0
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)
                     TAG_LIB=UI-R-Y0
                     TAG_TISSUE=Eye
                     TAG_SEQ=CATTG"
BASE COUNT      105 a     76 c     46 g     96 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 24;  Length 323;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      24 ATTAATGAAAAAAAAAAAAA 4


RESULT  34
AI698546/c
LOCUS       AI698546      330 bp    mRNA            EST       07-MAR-2000
DEFINITION  wa79c05.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2302376 3' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   AI698546
VERSION     AI698546.1  GI:4986446
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

```
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1386  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 307.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2302376"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       72 a      70 c      57 g     130 t       1 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 23;  Length 330;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    388 gaaaaaaaaaaaaaatgcttg 408
          |||||||||||||||||||||
Db    161 GAAAAAAAAAAAAAATGCTTG 141


RESULT  35
R84108
LOCUS       R84108         366 bp    mRNA            EST       04-AUG-1998
DEFINITION  16067 Lambda-PRL2 Arabidopsis thaliana cDNA clone 154C4T7, mRNA
            sequence.
ACCESSION   R84108
VERSION     R84108.1  GI:2763694
KEYWORDS    EST.
```



```
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Jan 9, 1998 this sequence version replaced gi:930542.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7 dye primer.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Arabidopsis thaliana"
                     /strain="var columbia"
                     /db_xref="taxon:3702"
                     /clone="154C4T7"
                     /clone_lib="Lambda-PRL2"
                     /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                     inserts were directionally cloned with Sal-Not arms using
                     oligo dT primed cDNA. "
BASE COUNT      106 a      87 c      86 g      73 t       14 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 144;  Length 366;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1800 aactgagattaaagatttcca 1820
           |||||||||||||||||||||
Db    213 AACTGAGATTAAAGATTTCCA 233
```



```
RESULT  36
AI060180/c
LOCUS       AI060180       372 bp    mRNA            EST       11-FEB-1999
DEFINITION  UI-R-C1-li-d-05-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
            UI-R-C1-li-d-05-0-UI 3', mRNA sequence.
ACCESSION   AI060180
VERSION     AI060180.1  GI:3333957
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 372)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 12-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .372
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-C1-li-d-05-0-UI"
                     /clone_lib="UI-R-C1"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
                     library is a subtracted library derived from the UI-R-C0
                     library, which is a subtracted library derived from the
                     UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
                     consisted of a mixture of  individually tagged normalized
                     libraries constructed from rat placenta,  adult lung,
                     brain, liver, kidney, heart, spleen, ovary, and muscle.
```



```
                     The UI-R-E1 library  consisted of a mixture of
                     individually tagged normalized libraries constructed from
                     8, 12 and 18-day embryo. The tag is a string of 3-5
                     nucleotides present between the Not 1 site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-C1) was constructed as follows: PCR
                     amplified cDNA inserts from  UI-R-C0 clones from which 3'
                     ESTs had been derived was used as a driver in a
                     hybridization with the UI-R-C0 library in the form of
                     single-stranded circles. The remaining single-stranded
                     circles (subtracted library) was purified by
                     hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-C1
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)."
BASE COUNT      111 a      69 c      71 g     120 t       1 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 15;  Length 372;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db     24 ATTAATGAAAAAAAAAAAAA 4


RESULT  37
AW048532/c
LOCUS       AW048532       374 bp    mRNA            EST       18-SEP-1999
DEFINITION  UI-M-BH1-alx-b-08-0-UI.s1 NIH_BMAP_M_S2 Mus musculus cDNA clone
            UI-M-BH1-alx-b-08-0-UI 3', mRNA sequence.
ACCESSION   AW048532
VERSION     AW048532.1  GI:5909061
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
```

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized amygdala library cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 60-95, >AT_rich#Low_complexity
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .374
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH1-alx-b-08-0-UI"
/clone_lib="NIH_BMAP_M_S2"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S2 library is a subtracted library derived from NIH_BMAP_M_S1, which in turn is a subtracted library derived from a mixture of normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH_BMAP_M_S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries.
TAG_LIB=NIH_BMAP_M_S2
TAG_TISSUE=amygdala
TAG_SEQ=GTGAG"
BASE COUNT 108 a 91 c 72 g 103 t
ORIGIN

Query Match 0.4%; Score 21; DB 39; Length 374;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 attaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||||
Db 24 ATTAATGAAAAAAAAAAAAAA 4

RESULT 38
AI969768/c
LOCUS AI969768 376 bp mRNA EST 08-MAR-2000
DEFINITION wq73f07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2476933 3', mRNA sequence.
ACCESSION AI969768
VERSION AI969768.1 GI:5766586
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 376)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 806 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 338.
FEATURES Location/Qualifiers
source 1. .376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2476933"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made

from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 119 a 67 c 63 g 127 t
ORIGIN

Query Match 0.4%; Score 21; DB 27; Length 376;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 attaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||||
Db 21 ATTAATGAAAAAAAAAAAAAA 1

RESULT 39
AI641714/c
LOCUS AI641714 377 bp mRNA EST 29-APR-1999
DEFINITION fc22h12.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA 3' similar to contains element LTR1 repetitive element ;, mRNA sequence.
ACCESSION AI641714
VERSION AI641714.1 GI:4720189
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE. 1 (bases 1 to 377)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:

www.rzpd.de)
zebrafish identity (p-value greater than 1e-99) found to: gi|2225224|gb|AA494796|AA494796 fa11a06.s1 Zebrafish ICRFzfls Danio rerio cDNA
Seq primer: T7 ET from Amersham
High quality sequence stop: 349.
FEATURES Location/Qualifiers
source 1. .377
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield stage embryos"
/lab_host="XL1-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3'}; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik,Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."
BASE COUNT 139 a 79 c 45 g 114 t
ORIGIN

Query Match 0.4%; Score 21; DB 23; Length 377;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 attaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||||
Db 21 ATTAATGAAAAAAAAAAAAAA 1

RESULT 40
BE227191
LOCUS BE227191 380 bp mRNA EST 06-JUL-2000
DEFINITION us76a05.y1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:3257264 5' similar to WP:ZK546.13 CE02913 ;, mRNA sequence.

```
ACCESSION   BE227191
VERSION     BE227191.1  GI:8932427
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: us78a05.xl
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1068628
            Seq primer: -40RP from Gibco
            High quality sequence stop: 378.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Mus musculus"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3257264"
                     /clone_lib="NCI_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and  Differentiation 7, 3-11 (1996)."
BASE COUNT      132 a      75 c      95 g      78 t
ORIGIN
```

Query Match 0.4%; Score 21; DB 105; Length 380;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
Qy   2403 tttagagaaagaagtagagaa 2423
          |||||||||||||||||||||
Db     84 TTTAGAGAAAGAAGTAGAGAA 104


RESULT  41
AL041690/c
LOCUS       AL041690      387 bp    mRNA            EST       29-FEB-2000
DEFINITION  DKFZp434N1917_s1 434 (synonym: htes3) Homo sapiens cDNA clone
            DKFZp434N1917 3', mRNA sequence.
ACCESSION   AL041690
VERSION     AL041690.1  GI:5421039
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Bloecker, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Bloecker H
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 3' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No rl sequence available.
            This clone (DKFZp434N1917) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .387
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434N1917"
                     /clone_lib="434 (synonym: htes3)"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT       63 a      104 c      103 g      117 t
ORIGIN
```

Query Match 0.4%; Score 21; DB 28; Length 387;

Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db     29 ATTAATGAAAAAAAAAAAAA 9


RESULT  42
AW528569/c
LOCUS       AW528569      391 bp    mRNA            EST       06-MAR-2000
DEFINITION  UI-R-BT1-ako-h-04-0-UI.s1 UI-R-BT1 Rattus norvegicus cDNA clone
            UI-R-BT1-ako-h-04-0-UI 3', mRNA sequence.
ACCESSION   AW528569
VERSION     AW528569.1  GI:7170983
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized corpus-striatum library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.reagen.com)
            Seq primer: M13 Forward
            POLYA="Yes.
FEATURES             Location/Qualifiers
     source          1. .391
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BT1-ako-h-04-0-UI"
                     /clone_lib="UI-R-BT1"
```



```
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The library
                     UI-R-BT1 is a subtracted library derived from a mixture of
                     the following tissues: hippocampus, thalamus, mid-brain,
                     medulla, corpus striatum, cerebral cortex and testis. For
                     a detailed description of the library from  which this
                     clone was derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB=UI-R-BT1
                     TAG_TISSUE=corpus-striatum
                     TAG_SEQ=CTAGG"
BASE COUNT       95 a      101 c      100 g       94 t       1 others
ORIGIN
```

Query Match 0.4%; Score 21; DB 91; Length 391;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db     24 ATTAATGAAAAAAAAAAAAA 4


RESULT  43
AA132821/c
LOCUS       AA132821      400 bp    mRNA            EST       27-NOV-1996
DEFINITION  zo18d04.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:587239 5', mRNA sequence.
ACCESSION   AA132821
VERSION     AA132821.1  GI:1694310
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
```

```
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 254.
FEATURES              Location/Qualifiers
     source           1. .400
                      /organism="Homo sapiens"
                      /db_xref="GDB:4615541"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:587239"
                      /clone_lib="Stratagene colon (#937204)"
                      /tissue_type="tumor"
                      /cell_line="T84 carcinoma cell line"
                      /lab_host="SOLR cells (kanamycin resistant)"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. T-84 colonic epithelial cell line.  Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                      CTCGACTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      138 a      56 c      51 g     151 t       4 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 2;  Length 400;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      38 ATTAATGAAAAAAAAAAAAA 18


RESULT  44
R63336/c
LOCUS       R63336         400 bp    mRNA            EST       26-MAY-1995
DEFINITION  yi07f09.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
            IMAGE:138569 5', mRNA sequence.
ACCESSION   R63336
VERSION     R63336.1  GI:835215
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 400)
```



```
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 628
            High qality sequence stops: 257
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 628    Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 257.
FEATURES              Location/Qualifiers
     source           1. .400
                      /organism="Homo sapiens"
                      /db_xref="GDB:544983"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:138569"
                      /clone_lib="Soares placenta Nb2HP"
                      /sex="Female"
                      /dev_stage="placenta obtained at birth (full term)"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      100 a      87 c     106 g      99 t       8 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 144;  Length 400;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttga 409
           |||||||||||||||||||||
Db      47 AAAAAAAAAAAAAATGCTTGA 27
```



```
RESULT  45
AI117620/c
LOCUS       AI117620       401 bp    mRNA            EST       02-SEP-1998
DEFINITION  ui49b02.x1 Sugano mouse embryo mewa Mus musculus cDNA clone
            IMAGE:1885611 3', mRNA sequence.
ACCESSION   AI117620
VERSION     AI117620.1  GI:3517944
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:969935
            Possible reversed clone: polyT not found
            Seq primer: custom primer used.
FEATURES              Location/Qualifiers
     source           1. .401
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1885611"
                      /clone_lib="Sugano mouse embryo mewa"
                      /dev_stage="embryo, 14 dpc"
                      /lab_host="DH10B"
                      /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                      Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                      with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                      double-stranded cDNA was ligated to a  DraIII adaptor
                      [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                      sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                      CACCATGTG). XhoI should be used to isolate the cDNA
                      insert.  Size selection was performed to exclude fragments
                      <1.5kb.  Library constructed by Dr. Sumio Sugano
```



```
                      (University of Tokyo Institute of Medical Science).
                      Custom primers  for sequencing: 5' end primer
                      CTTCTGCTCTAAAAGCTGCG and 3' end primer
                      CGACCTGCAGCTCGAGCACA."
BASE COUNT       88 a       89 c      100 g      124 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 16;  Length 401;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      28 ATTAATGAAAAAAAAAAAAA 8


RESULT  46
AA505862/c
LOCUS       AA505862       407 bp    mRNA            EST       20-AUG-1997
DEFINITION  nh99e04.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:966654 3',
            mRNA sequence.
ACCESSION   AA505862
VERSION     AA505862.1  GI:2241999
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph:D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 271    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 344.
FEATURES              Location/Qualifiers
     source           1. .407
                      /organism="Homo sapiens"
```

/db_xref="taxon:9606"
/clone="IMAGE:966654"
/clone_lib="NCI_CGAP_Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. This library is the normalized version of NCI_CGAP_Br1.1. Library was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 148 a 59 c 55 g 145 t
ORIGIN

Query Match 0.4%; Score 21; DB 8; Length 407;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 attaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 22 ATTAATGAAAAAAAAAAAAAA 2

RESULT 47
AI842076/c
LOCUS AI842076 408 bp mRNA EST 14-JUL-1999
DEFINITION UI-M-AN1-afg-b-10-0-UI.s1 NIH_BMAP_MBG_N Mus musculus cDNA clone. UI-M-AN1-afg-b-10-0-UI 3', mRNA sequence.
ACCESSION AI842076
VERSION AI842076.1 GI:5476289
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 408)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706

Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized basal ganglia library cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .408
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AN1-afg-b-10-0-UI"
/clone_lib="NIH_BMAP_MBG_N"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MBG_N library is a normalized library constructed from mouse basal ganglia. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806 , 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
TAG_LIB=NIH_BMAP_MBG_N
TAG_TISSUE=basal-ganglia
TAG_SEQ=TGTAC"
BASE COUNT 120 a 104 c 73 g 111 t
ORIGIN

Query Match 0.4%; Score 21; DB 25; Length 408;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 attaatgaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 24 ATTAATGAAAAAAAAAAAAA 4

RESULT 48
AL377900
LOCUS AL377900 409 bp mRNA EST 03-AUG-2000

DEFINITION MtBB34F10F1 MtBB Medicago truncatula cDNA clone MtBB34F10 T3, mRNA sequence.
ACCESSION AL377900
VERSION AL377900.1 GI:9677652
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 409)
AUTHORS Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson ,V. and Gamas,P.
TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
JOURNAL Unpublished (2000)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES Location/Qualifiers
source 1. .409
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBB34F10"
/clone_lib="MtBB"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
/note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."
BASE COUNT 135 a 68 c 83 g 123 t

ORIGIN

Query Match 0.4%; Score 21; DB 28; Length 409;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 attaatgaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 388 ATTAATGAAAAAAAAAAAAA 408

RESULT 49
D76720
LOCUS D76720 410 bp mRNA EST 07-OCT-1996
DEFINITION MUS69B04 mouse embryonal carcinoma cell line F9 Mus musculus cDNA clone 69B04, mRNA sequence.
ACCESSION D76720
VERSION D76720.1 GI:1596249
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 410)
AUTHORS Nishiguchi,S., Sakuma,R., Nomura,M., Zou,Z., Jeeranaisilavong,J., Joh,T., Yasunaga,T. and Shimada,K.
TITLE A catalogue of genes in mouse embryonal carcinoma F9 cells identified with expressed sequence tags
JOURNAL J. Biochem. (Tokyo) 119 (4), 749-767 (1996)
MEDLINE 96337530
COMMENT Contact: Kazunori Shimada
Department of Medical Genetics, Division of Molecular Biomedicine
Research Institute for Microbial Diseases, Osaka University
3-1, Yamadaoka, Suita, Osaka, 565, Japan
Tel: 06-879-8325
Fax: 06-879-8326.
FEATURES Location/Qualifiers
source 1. .410
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="69B04"
/clone_lib="mouse embryonal carcinoma cell line F9"
BASE COUNT 121 a 73 c 127 g 81 t 8 others
ORIGIN

Query Match 0.4%; Score 21; DB 140; Length 410;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   2403 tttagagaaaagaagtagagaa 2423
          |||||||||||||||||||||||
Db    310 TTTAGAGAAAGAAGTAGAGAA 330


RESULT  50
BE951351
LOCUS       BE951351      413 bp    mRNA            EST       04-OCT-2000
DEFINITION  UI-M-CC0-axz-g-02-0-UI.s1 NIH_BMAP_Ret1 Mus musculus cDNA clone
            UI-M-CC0-axz-g-02-0-UI 3', mRNA sequence.
ACCESSION   BE951351
VERSION     BE951351.1  GI:10590696
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            embryonic retina tissue cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: Researchers may obtain BMAP cDNA clones from
            RESEARCH GENETICS. It should be noted that Bento Soares is
            generating a small number of additional specialized non-redundant
            arrays of BMAP cDNAs whose availability will be considered under
            appropriate and limited collaborative arrangements The following
            repetitive elements were found in this cDNA sequence: 29-74,
            >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-CC0-axz-g-02-0-UI"
```



```
                     /clone_lib="NIH_BMAP_Ret1"
                     /dev_stage="13.5 days pc"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret1 library is derived from mouse embryonic
                     retina tissue. For a detailed description of the library
                     from which this clone was derived, please visit our web
                     site at brainest.eng.uiowa.edu.
                     TAG_LIB=NIH_BMAP_Ret1
                     TAG_TISSUE=embryonic-retina
                     TAG_SEQ=CTGCTAGTAGCA"
BASE COUNT      122 a      79 c     110 g     101 t       1 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 138;  Length 413;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    386 atgaaaaaaaaaaaaaatgct 406
          |||||||||||||||||||||
Db    383 ATGAAAAAAAAAAAAATGCT 403


RESULT  51
R33351/c
LOCUS       R33351        415 bp    mRNA            EST       28-APR-1995
DEFINITION  yh77f10.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
            IMAGE:135787 5', mRNA sequence.
ACCESSION   R33351
VERSION     R33351.1  GI:789209
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 584
```



```
            High quality sequence stops: 251
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 584    Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 251.
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Homo sapiens"
                     /db_xref="GDB:541762"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:135787"
                     /clone_lib="Soares placenta Nb2HP"
                     /sex="Female"
                     /dev_stage="placenta obtained at birth (full term)"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT       99 a      89 c     113 g     108 t       6 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 144;  Length 415;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaatgcttga 409
          |||||||||||||||||||||
Db     47 AAAAAAAAAAAAATGCTTGA 27


RESULT  52
AQ198451
LOCUS       AQ198451      419 bp    DNA             GSS       20-APR-1999
DEFINITION  RPCI11-48M22.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-48M22,
            DNA sequence.
ACCESSION   AQ198451
VERSION     AQ198451.1  GI:3610650
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```



```
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-48M22.TK
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Homo sapiens"
                     /db_xref="GDB:7518357"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-48M22"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      161 a      78 c      60 g     120 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 150;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaat 403
          |||||||||||||||||||||
Db    181 TTAATGAAAAAAAAAAAAAT 201


RESULT  53
AW321430
LOCUS       AW321430      423 bp    mRNA            EST       25-JAN-2000
DEFINITION  uo31g02.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:2631122 5'
            similar to WP:ZK546.13 CE02913 ;, mRNA sequence.
ACCESSION   AW321430
VERSION     AW321430.1  GI:6750974
```

```
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            MGI:1024630
            Seq primer: -40RP from Gibco
            High quality sequence stop: 266.
FEATURES             Location/Qualifiers
     source          1..423
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2631122"
                     /clone_lib="NCI_CGAP_Mam6"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      140 a      88 c     107 g      87 t       1 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 89;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2403 tttagagaaagaagtagagaa 2423
           |||||||||||||||||||||
Db     84 TTTAGAGAAAGAAGTAGAGAA 104
```



```
RESULT  54
AA855552
LOCUS       AA855552      428 bp    mRNA            EST       06-MAR-1998
DEFINITION  vw74f12.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:1260719 5' similar to WP:ZK546.13 CE02913 ;, mRNA sequence.
ACCESSION   AA855552
VERSION     AA855552.1  GI:2943090
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:663271
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 417.
FEATURES             Location/Qualifiers
     source          1..428
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1260719"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 91 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; 5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      143 a      94 c     110 g      81 t
```



```
ORIGIN


  Query Match             0.4%;  Score 21;  DB 12;  Length 428;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2403 tttagagaaagaagtagagaa 2423
           |||||||||||||||||||||
Db     82 TTTAGAGAAAGAAGTAGAGAA 102


RESULT  55
AL377901
LOCUS       AL377901      430 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBB34F10R1 MtBB Medicago truncatula cDNA clone MtBB34F10 T7, mRNA
            sequence.
ACCESSION   AL377901
VERSION     AL377901.1  GI:9677653
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
            nodules
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1..430
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBB34P10"
                     /clone_lib="MtBB"
                     /tissue_type="symbiotic root nodules"
                     /dev_stage="harvested 4 days post inoculation with
                     Sinorhizobium meliloti"
```



```
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber on
                     nitrogen-rich medium for 21 days. Three days before
                     inoculation with Sinorhizobium meliloti, the medium was
                     replaced by N-free medium. Root nodules (+ short adjacent
                     root segments) were harvested 4 days post inoculation.
                     cDNA was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zap XR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT      138 a      74 c      93 g     125 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db    388 ATTAATGAAAAAAAAAAAAAA 408


RESULT  56
BE756820
LOCUS       BE756820      431 bp    mRNA            EST       15-SEP-2000
DEFINITION  211079 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE756820
VERSION     BE756820.1  GI:10170812
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
```

```
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 62  row: D  column: 6
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .431
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 2BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from testis, thymus,
                      semitendonosus muscle, longissimus muscle, pancreas,
                      adrenal, and endometrium."
BASE COUNT      136 a      72 c      75 g     148 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 135;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttga 409
           ||||||||||||||||||||
Db     113 AAAAAAAAAAAAAATGCTTGA 133


RESULT  57
BE508093/c
LOCUS       BE508093      436 bp    mRNA            EST       07-AUG-2000
DEFINITION  dc26e02.y1 NICHD XGC Lil Xenopus laevis cDNA clone IMAGE:3398234
            5', mRNA sequence.
ACCESSION   BE508093
VERSION     BE508093.1  GI:9727868
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 436)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```



```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 435.
FEATURES              Location/Qualifiers
     source           1. .436
                      /organism="Xenopus laevis"
                      /db_xref="taxon:8355"
                      /clone="IMAGE:3398234"
                      /clone_lib="NICHD XGC Lil"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                      Average insert size 1.4 kb. Constructed by Life
                      Technologies. Note: This is a Xenopus Gene Collection (XGC
                      ) library."
BASE COUNT      126 a      80 c      84 g     145 t       1 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 109;  Length 436;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     214 ATTAATGAAAAAAAAAAAAAA 194


RESULT  58
CNS00QEZ/c
LOCUS       CNS00QEZ      438 bp    DNA             GSS       28-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence SP6 end of BAC F11E24
            of IGF library from strain Columbia of Arabidopsis thaliana,
            genomic survey sequence.
ACCESSION   AL085721
VERSION     AL085721.1  GI:5286861
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
```



```
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 438)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES              Location/Qualifiers
     source           1. .438
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /clone_lib="IGF"
                      /clone="F11E24"
                      /note="end : SP6"
BASE COUNT      107 a      79 c     100 g     152 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 190;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1800 aactgagattaaagatttcca 1820
           ||||||||||||||||||||
Db     140 AACTGAGATTAAAGATTTCCA 120


RESULT  59
AI911383/c
LOCUS       AI911383      439 bp    mRNA            EST       17-DEC-1999
DEFINITION  wd16e02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2328314 3', mRNA sequence.
ACCESSION   AI911383
VERSION     AI911383.1  GI:5631119
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
```



```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 551  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 435.
FEATURES              Location/Qualifiers
     source           1. .439
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2328314"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo. "
BASE COUNT      124 a      100 c     100 g     115 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 26;  Length 439;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      22 ATTAATGAAAAAAAAAAAAAA 2


RESULT  60
AL368910/c
LOCUS       AL368910      440 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBA27F02F1 MtBA Medicago truncatula cDNA clone MtBA27F02 T3, mRNA
            sequence.
ACCESSION   AL368910
VERSION     AL368910.1  GI:9668663
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 440)
```

```
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from nitrogen-starved roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1..440
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBA27F02"
                     /clone_lib="MtBA"
                     /tissue_type="root tips"
                     /dev_stage="harvested after 3 days of N-starvation"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber for 14
                     days on nitrogen-rich medium followed by 3 days on N-free
                     medium. RNA was extracted from root tips (1-3 cm). cDNA
                     was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zapXR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT      154 a      92 c      66 g     128 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 440;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      32 ATTAATGAAAAAAAAAAAAA 12


RESULT  61
AI425943/c
LOCUS       AI425943      442 bp    mRNA            EST       09-MAR-1999
```



```
DEFINITION  mg27d07.x1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:425005 3', mRNA sequence.
ACCESSION   AI425943
VERSION     AI425943.1  GI:4271874
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 402.
FEATURES             Location/Qualifiers
     source          1..442
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:425005"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo. "
```



```
BASE COUNT      158 a      66 c      72 g     146 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 20;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      23 ATTAATGAAAAAAAAAAAAA 3


RESULT  62
BE954247/c
LOCUS       BE954247      448 bp    mRNA            EST       04-OCT-2000
DEFINITION  UI-M-CC1-bad-f-07-0-UI.s1 NIH_BMAP_Ret1_N Mus musculus cDNA clone
            UI-M-CC1-bad-f-07-0-UI 3', mRNA sequence.
ACCESSION   BE954247
VERSION     BE954247.1  GI:10596577
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            embryonic retina tissue cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: Researchers may obtain BMAP cDNA clones from
            RESEARCH GENETICS. It should be noted that Bento Soares is
            generating a small number of additional specialized non-redundant
            arrays of BMAP cDNAs whose availability will be considered under
            appropriate and limited collaborative arrangements The following
            repetitive elements were found in this cDNA sequence: 60-95,
            >AT_rich#Low_complexity
```



```
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..448
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-CC1-bad-f-07-0-UI"
                     /clone_lib="NIH_BMAP_Ret1_N"
                     /dev_stage="13.5 days pc"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret1_N library is a normalized library derived
                     NIH_BMAP_Ret1. NIH_BMAP_Ret1 was made from  mouse
                     embryonic retina tissue. For a detailed description of the
                     library from which this clone was  derived, please visit
                     our web site at brainest.eng.uiowa.edu.
                     TAG_LIB=NIH_BMAP_Ret1_N
                     TAG_TISSUE=embryonic-retina
                     TAG_SEQ=CTGCTAGTAGCA"
BASE COUNT      126 a     109 c      96 g     117 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 138;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      24 ATTAATGAAAAAAAAAAAAA 4


RESULT  63
AQ111049/c
LOCUS       AQ111049      450 bp    DNA             GSS       29-AUG-1998
DEFINITION  CIT-HSP-2377I18.TR CIT-HSP Homo sapiens genomic clone 2377I18, DNA
            sequence.
ACCESSION   AQ111049
VERSION     AQ111049.1  GI:3487739
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
```

```
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2377I18.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301-838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..450
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2377I18"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      140 a      67 c      70 g     173 t
ORIGIN


  Query Match           0.4%;  Score 21;  DB 149;  Length 450;
  Best Local Similarity  100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3170 tcaataacaataacaataatg 3190
          |||||||||||||||||||||
Db     95 TCAATAACAATAACAATAATG 75


RESULT  64
AI715886/c
LOCUS       AI715886      454 bp    mRNA            EST       10-JUN-1999
DEFINITION  UI-R-Y0-abv-e-06-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-abv-e-06-0-UI 3', mRNA sequence.
ACCESSION   AI715886
VERSION     AI715886.1  GI:5033139
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
```



```
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized Eye library cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.resgen.com) The following repetitive elements were
            found in this cDNA sequence: 60-95, >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..454
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-Y0-abv-e-06-0-UI"
                     /clone_lib="UI-R-Y0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
                     library is a subtracted library derived from an
                     individually-tagged normalized whole-eye (minus the lens)
                     library. The driver for the subtraction consisted of a
                     pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
                     UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
                     3-5 nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-Y0) was constructed as follows: PCR
                     amplified cDNA inserts from previous library clones from
                     which 3' ESTs had been derived were used as a driver in a
                     hybridization with the normalized whole-eye library in
                     the form of single-stranded circles. The remaining
                     single-stranded circles (subtracted library) was purified
                     by hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
```



```
                     bacteria (Life Technologies) to generate the UI-R-Y0
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)
                     TAG_LIB=UI-R-Y0
                     TAG_TISSUE=Eye
                     TAG_SEQ=CATTG"
BASE COUNT      133 a     101 c     103 g     117 t
ORIGIN


  Query Match           0.4%;  Score 21;  DB 24;  Length 454;
  Best Local Similarity  100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db     24 ATTAATGAAAAAAAAAAAAA 4


RESULT  65
AZ395583/c
LOCUS       AZ395583      454 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0159F18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0159F18 R, DNA sequence.
ACCESSION   AZ395583
VERSION     AZ395583.1  GI:10510655
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0159  row: F  column: 18
            Seq primer: CACACAGGAAACAGCTATGACC
```



```
            Class: plasmid ends
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1..454
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0159F18"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      126 a     117 c      91 g     120 t
ORIGIN


  Query Match           0.4%;  Score 21;  DB 180;  Length 454;
  Best Local Similarity  100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    141 gtaagcatgacacacaaacca 161
          |||||||||||||||||||||
Db    192 GTAAGCATGACACACAAACCA 172


RESULT  66
AW074370/c
LOCUS       AW074370      457 bp    mRNA            EST       13-OCT-1999
DEFINITION  xb11h04.x1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2576023 3',
            mRNA sequence.
ACCESSION   AW074370
VERSION     AW074370.1  GI:6029368
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2576023"
                     /clone_lib="NCI_CGAP_GU1"
                     /tissue_type="2 pooled high-grade transitional cell
                     tumors"
                     /lab_host="DH10B"
                     /note="Organ: genitourinary tract; Vector: pCMV-SPORT6;
                     Site_1: SalI; Site_2: NotI; Cloned unidirectionally.
                     Primer: Oligo dT.   Library constructed by Life
                     Technologies."
BASE COUNT      169 a      71 c      61 g     156 t
ORIGIN

  Query Match             0.4%;  Score 21;  DB 39;  Length 457;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      21 ATTAATGAAAAAAAAAAAAA 1


RESULT  67
AI378562/c
LOCUS       AI378562      460 bp    mRNA            EST       18-MAR-1999
```



```
DEFINITION  tc72a10.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:2070138
            3', mRNA sequence.
ACCESSION   AI378562
VERSION     AI378562.1  GI:4188415
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 824   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2070138"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      122 a      97 c      94 g     146 t       1 others
ORIGIN

  Query Match             0.4%;  Score 21;  DB 19;  Length 460;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
```



```
           |||||||||||||||||||||
Db      28 ATTAATGAAAAAAAAAAAAA 8


RESULT  68
BE943917/c
LOCUS       BE943917      461 bp    mRNA            EST       03-OCT-2000
DEFINITION  UI-M-BH3-avo-g-09-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-avo-g-09-0-UI 3', mRNA sequence.
ACCESSION   BE943917
VERSION     BE943917.1  GI:10521676
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            olfactory bulbs tissue cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: Researchers may obtain BMAP cDNA clones from
            RESEARCH GENETICS. It should be noted that Bento Soares is
            generating a small number of additional specialized non-redundant
            arrays of BMAP cDNAs whose availability will be considered under
            appropriate and limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-avo-g-09-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
```



```
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately  derived from a mixture of individually tagged
                     normalized libraries from  ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory  bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland,  striatum,
                     hipoccampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1.  The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows:  PCRamplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as  a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of  single-stranded circles.  The
                     remaining single-stranded circles  (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B  bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S4 library.  This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB=NIH_BMAP_M_S4
                     TAG_TISSUE=olfactory-bulbs
                     TAG_SEQ=CATGG"
BASE COUNT      128 a      113 c      71 g     149 t
ORIGIN

  Query Match             0.4%;  Score 21;  DB 137;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      24 ATTAATGAAAAAAAAAAAAA 4


RESULT  69
AI071557/c
LOCUS       AI071557      463 bp    mRNA            EST       11-FEB-1999
DEFINITION  UI-R-C1-ky-d-12-0-UI.s2 UI-R-C1 Rattus norvegicus cDNA clone
            UI-R-C1-ky-d-12-0-UI 3', mRNA sequence.
ACCESSION   AI071557
VERSION     AI071557.1  GI:3397772
KEYWORDS    EST.
SOURCE      Norway rat.
```

```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 12-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .463
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-C1-ky-d-12-0-UI"
                      /clone_lib="UI-R-C1"
                      /dev_stage="Adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
                      library is a subtracted library derived from the UI-R-C0
                      library, which is a subtracted library derived from the
                      UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
                      consisted of a mixture of individually tagged normalized
                      libraries constructed from rat placenta, adult lung,
                      brain, liver, kidney, heart, spleen, ovary, and muscle.
                      The UI-R-E1 library consisted of a mixture of
                      individually tagged normalized libraries constructed from
                      8, 12 and 18-day embryo. The tag is a string of 3-5
                      nucleotides present between the Not I site and the
                      oligo-dT track which allows identification of the library
                      of origin of a clone within the mixture. The subtracted
                      library (UI-R-C1) was constructed as follows: PCR
                      amplified cDNA inserts from UI-R-C0 clones from which 3'
                      ESTs had been derived was used as a driver in a
                      hybridization with the UI-R-C0 library in the form of
                      single-stranded circles. The remaining single-stranded
```



```
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-C1
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
BASE COUNT      125 a    102 c     90 g    145 t      1 others
ORIGIN
```

```
  Query Match             0.4%;  Score 21;  DB 15;  Length 463;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||||
Db      24 ATTAATGAAAAAAAAAAAAAA 4


RESULT  70
AW297911/c
LOCUS       AW297911       464 bp    mRNA            EST       16-JAN-2000
DEFINITION  UI-H-BW0-ajn-a-07-0-UI.s1 NCI_CGAP_Sub6 Homo sapiens cDNA clone
            IMAGE:2732077 3', mRNA sequence.
ACCESSION   AW297911
VERSION     AW297911.1  GI:6704547
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html The following repetitive
            elements were found in this cDNA sequence: 1-35,
            >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
```



```
FEATURES              Location/Qualifiers
     source           1. .464
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2732077"
                      /clone_lib="NCI_CGAP_Sub6"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub6
                      is a subtracted library derived from BW, which consists of
                      a mixture of four normalized libraries: NCI_CGAP_Brn50,
                      NCI_CGAP_Lul3, NCI_CGAP_Ov18, GBC1 . The NCI_CGAP_Sub6
                      library had 7 million recombinants. A single-stranded DNA
                      preparation of BW was used as a tracer in a subtractive
                      hybridization with a driver comprising: the IMAGE pool
                      (NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
                      3798-3803 (IMAGE CloneIDs 1322376-1323911,
                      1456008-1456775,1500552-1502855); NCI_CGAP_Kid5 pool 1
                      LLAM 3338-3342,3722-3725, 3776-3778 (IMAGE CloneIDs
                      1323912-1325831, 1471368-1472903, 1492104-1493255);
                      NCI_CGAP_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
                      CloneIDs 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
                      pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
                      CloneIDs 1257096-1258631,1469064-1470983, 1475592-1476743
                      ); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759,
                      3062-3068 (IMAGE CloneIDs 985608-986759, 1101192-1101959,
                      1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,
                      2871-2872 (IMAGE CloneIDs 1057416-1061255,1144584-1145351
                      ). (50% of the driver population), plus a pool of 3,840
                      arrayed clones from NCI_CGAP_Sub1 (IMAGE CloneIDs
                      2708616-2710535) and NCI_CGAP_Sub2 (IMAGE CloneIDs
                      2710536-2712455) (20% of the driver population), plus a
                      pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE CloneIDs
                      2712456-2723591) (30% of the driver population).
                      Subtraction was performed as previously described [Bonaldo
                      , Lennon & Soares (1996): Normalization and Subtraction:
                      Two Approaches To Facilitate Gene Discovery. Genome
                      Research 6, 791-806.
                      TAG_LIB=Gbc1
                      TAG_TISSUE=B cells germinal
                      TAG_SEQ=TCA"
BASE COUNT      177 a     75 c     47 g    165 t
ORIGIN
```

```
  Query Match             0.4%;  Score 21;  DB 88;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||||
```



```
Db      24 ATTAATGAAAAAAAAAAAAAA 4


RESULT  71
AA533164/c
LOCUS       AA533164       470 bp    mRNA            EST       21-AUG-1997
DEFINITION  nj60f12.s1 NCI_CGAP_Pr9 Homo sapiens cDNA clone IMAGE:996911
            similar to gb:X70476 COATOMER BETA' SUBUNIT (HUMAN);, mRNA
            sequence.
ACCESSION   AA533164
VERSION     AA533164.1  GI:2277260
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 684  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 316.
FEATURES              Location/Qualifiers
     source           1. .470
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:996911"
                      /clone_lib="NCI_CGAP_Pr9"
                      /sex="male"
                      /tissue_type="normal prostatic epithelial cells"
                      /lab_host="DH10B"
                      /note="Organ: prostate; Vector: pAMP10; mRNA made from
                      normal prostatic epithelial cells, cDNA made by oligo-dT
                      priming. Non-directionally cloned. Size-selected on
                      agarose gel, average insert size 600 bp. Library made by
                      D. Krizman, NIH."
BASE COUNT      140 a     87 c     83 g    160 t
ORIGIN
```

```
  Query Match             0.4%;  Score 21;  DB 8;  Length 470;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      28 ATTAATGAAAAAAAAAAAAA 8


RESULT  72
BE099767/c
LOCUS       BE099767      479 bp    mRNA            EST       13-JUN-2000
DEFINITION  UI-R-BJ1-atq-a-07-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
            UI-R-BJ1-atq-a-07-0-UI 3', mRNA sequence.
ACCESSION   BE099767
VERSION     BE099767.1  GI:8491647
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized AV canal at 15 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
```



```
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ1-atq-a-07-0-UI"
                     /clone_lib="UI-R-BJ1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
                     library is a subtracted library derived from the following
                     tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
                     canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
                     AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
                     For a detailed description of the library from which this
                     clone was derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB=UI-R-BJ1
                     TAG_TISSUE=AV canal at 15 dpc
                     TAG_SEQ=GAAGG"
BASE COUNT      113 a     89 c    105 g    172 t
ORIGIN

  Query Match             0.4%;  Score 21;  DB 134;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      24 ATTAATGAAAAAAAAAAAAA 4


RESULT  73
AA964331/c
LOCUS       AA964331      480 bp    mRNA            EST       09-MAR-1999
DEFINITION  UI-R-C0-gu-g-12-0-UI.s1 UI-R-C0 Rattus norvegicus cDNA clone
            UI-R-C0-gu-g-12-0-UI 3', mRNA sequence.
ACCESSION   AA964331
VERSION     AA964331.1  GI:4279206
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On May 18, 1998 this sequence version replaced gi:3137823.
```



```
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 18-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-C0-gu-g-12-0-UI"
                     /clone_lib="UI-R-C0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
                     library is a subtracted library derived from the UI-R-A1
                     and UI-R-E1 libraries. The UI-R-A1 library consisted of a
                     mixture of individually tagged normalized libraries
                     constructed from rat  placenta, adult lung, brain, liver,
                     kidney, heart, spleen, ovary,  and muscle. The UI-R-E1
                     library  consisted of a mixture of individually tagged
                     normalized libraries constructed from 9, 12 and 18-day
                     embryo.  The tag is a string of 3-5 nucleotides present
                     between the Not I  site and the oligo-dT track which
                     allows identification of the library of origin of a clone
                     within the mixture. The subtracted library (UI-R-C0) was
                     constructed as follows: PCR amplified cDNA inserts from a
                     pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
                     been derived was used as a driver in a hybridization with
                     the pooled UI-R-A1 and UI-R-E1 library in the form of
                     single-stranded circles.  The remaining single-stranded
                     circles (subtracted library) was purified by
                     hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-C0
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)"
BASE COUNT      113 a     89 c    106 g    172 t
ORIGIN
```



```
  Query Match             0.4%;  Score 21;  DB 14;  Length 480;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      24 ATTAATGAAAAAAAAAAAAA 4


RESULT  74
AT002267/c
LOCUS       AT002267      492 bp    mRNA            EST       25-AUG-1999
DEFINITION  AT002267 Antheraea yamamai posterior silkgland Antheraea yamamai
            cDNA clone AY571, mRNA sequence.
ACCESSION   AT002267
VERSION     AT002267.1  GI:5771132
KEYWORDS    EST.
SOURCE      Japanese oak silkmoth.
  ORGANISM  Antheraea yamamai
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
            ; Bombycoidea; Saturniidae; Saturniinae; Antheraea.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Yun,E.Y., Kang,S.W., Lee,H.W., Hwang,J.S. and Kim,K.Y.
  TITLE     Antheraea yamamai EST
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yun EY
            Department of Sericulture and Entomology
            National Institute of Agricultural Science and Technology
            RDA 61 Seodundong, Suwon, Kyung-gi 441 - 100, Korea
            Tel: 82-331-290-8546
            Fax: 82-331-290-8420
            Email: yuneye@niast.go.kr
            Submitted through BRIC(Biological Research Information Center) of
            Korea
            URL: http://bric.postech.ac.kr/.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Antheraea yamamai"
                     /db_xref="taxon:7121"
                     /clone="AY571"
                     /clone_lib="Antheraea yamamai posterior silkgland"
                     /lab_host="XL1-Blue MRF', SOLR"
                     /note="Vector: Uni-ZAP XR; Site_1: EcoRI; Site_2: XhoI;
                     average insert size:1000 bp;initial pfu:1.100 X 10 6"
BASE COUNT       92 a    143 c    130 g    127 t
ORIGIN

  Query Match             0.4%;  Score 21;  DB 28;  Length 492;
  Best Local Similarity   100.0%;  Pred. No. 71;
```

```
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||||
Db      24 ATTAATGAAAAAAAAAAAAAA 4


RESULT  75
AL374239
LOCUS       AL374239      497 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBB05E01R1 MtBB Medicago truncatula cDNA clone MtBB05E01 T7, mRNA
            sequence.
ACCESSION   AL374239
VERSION     AL374239.1  GI:9673991
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
            nodules
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBB05E01"
                     /clone_lib="MtBB"
                     /tissue_type="symbiotic root nodules"
                     /dev_stage="harvested 4 days post inoculation with
                     Sinorhizobium meliloti"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber on
                     nitrogen-rich medium for 21 days. Three days before
                     inoculation with Sinorhizobium meliloti, the medium was
                     replaced by N-free medium. Root nodules (+ short adjacent
```



```
                     root segments) were harvested 4 days post inoculation.
                     cDNA was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zap XR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT      147 a      73 c     125 g     152 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 497;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||||
Db     473 ATTAATGAAAAAAAAAAAAAA 493


RESULT  76
AL374022
LOCUS       AL374022      500 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBB04C01R1 MtBB Medicago truncatula cDNA clone MtBB04C01 T7, mRNA
            sequence.
ACCESSION   AL374022
VERSION     AL374022.1  GI:9673774
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
            nodules
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
```



```
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBB04C01"
                     /clone_lib="MtBB"
                     /tissue_type="symbiotic root nodules"
                     /dev_stage="harvested 4 days post inoculation with
                     Sinorhizobium meliloti"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber on
                     nitrogen-rich medium for 21 days. Three days before
                     inoculation with Sinorhizobium meliloti, the medium was
                     replaced by N-free medium. Root nodules (+ short adjacent
                     root segments) were harvested 4 days post inoculation.
                     cDNA was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zap XR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT      157 a      64 c     126 g     153 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 500;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||||
Db     466 ATTAATGAAAAAAAAAAAAAA 486


RESULT  77
BE809028
LOCUS       BE809028      509 bp    mRNA            EST       20-SEP-2000
DEFINITION  214367 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE809028
VERSION     BE809028.1  GI:10240140
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
```



```
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 70  row: H  column: 14
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT      165 a      93 c      82 g     169 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 136;  Length 509;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||||
Db     482 ATTAATGAAAAAAAAAAAAAA 502


RESULT  78
AQ218828
LOCUS       AQ218828      511 bp    DNA             GSS       19-SEP-1998
DEFINITION  HS_3248_B1_D11_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3248 Col=21 Row=H, DNA sequence.
ACCESSION   AQ218828
VERSION     AQ218828.1  GI:3630455
KEYWORDS    GSS.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3248  row: H  column: 21
            Class: BAC ends
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3248 Col=21 Row=H"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      158 a    103 c    125 g    125 t
ORIGIN


 Query Match             0.4%;  Score 21;  DB 150;  Length 511;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgcttga 409
          |||||||||||||||||||||
Db    406 AAAAAAAAAAAAAATGCTTGA 426


RESULT  79
AL385814
LOCUS       AL385814      513 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBC30G12R1 MtBC Medicago truncatula cDNA clone MtBC30G12 T7, mRNA
            sequence.
ACCESSION   AL385814
```



```
VERSION     AL385814.1  GI:9685565
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from endomycorrhizal roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBC30G12"
                     /clone_lib="MtBC"
                     /tissue_type="arbuscular mycorrhiza"
                     /dev_stage="harvested 3 weeks post inoculation with Glomus
                     intraradices"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; M. truncatula sterilised seeds were germinated for
                     72h at 25 C, before transplanting into a 1/3 Epoisses soil
                     : 2/3 calcined Terragreen mix in the presence of onion
                     root fragments colonized by the arbuscular mycorrhizal
                     fungus Glomus intraradices (Schenck & Smith, isolate LPA8
                     ). The plants were watered every day and twice a week with
                     a modified nutrient Long Ashton solution without phosphate
                     but with a high level of nitrate. After 3 weeks RNA was
                     extracted from whole root systems. cDNA was prepared from
                     polyA+ enriched RNA. The cDNA was directionally ligated
                     into Uni-zap XR vector from Stratagene and packaged using
                     Gigapack Gold packaging extracts. Plasmids containing cDNA
                     inserts were mass-excised from phage stocks using ExAssit
                     helper phage and propagated in SOLR cells. Clone ordering
                     and sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France). Note : EST may be of
                     fungal origin."
BASE COUNT      162 a     78 c    123 g    150 t
```



```
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 513;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db    482 ATTAATGAAAAAAAAAAAAA 502


RESULT  80
AU096044
LOCUS       AU096044      519 bp    mRNA            EST       30-JUN-2000
DEFINITION  AU096044 Rice green shoot Oryza sativa cDNA clone S11953, mRNA
            sequence.
ACCESSION   AU096044
VERSION     AU096044.1  GI:8858726
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from green shoot (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
            S11953_10Z.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="S11953"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT      136 a    147 c    121 g    113 t      2 others
ORIGIN
```



```
  Query Match             0.4%;  Score 21;  DB 29;  Length 519;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaa 402
          |||||||||||||||||||||
Db    483 ATTAATGAAAAAAAAAAAAA 503


RESULT  81
AA832599/c
LOCUS       AA832599      522 bp    mRNA            EST       23-FEB-1998
DEFINITION  vw43f03.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1246589 5', mRNA sequence.
ACCESSION   AA832599
VERSION     AA832599.1  GI:2906327
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:660277
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1246589"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
```

```
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker: Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      94 a    157 c    103 g    168 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 12;  Length 522;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaat 403
           |||||||||||||||||||||
Db     147 TTAATGAAAAAAAAAAAAAT 127


RESULT  82
BE033628
LOCUS       BE033628      533 bp    mRNA            EST       07-JUN-2000
DEFINITION  MF05C02 MF Mesembryanthemum crystallinum cDNA 5', mRNA sequence.
ACCESSION   BE033628
VERSION     BE033628.1  GI:8328637
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Perrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            No hits found using blastx vs. genbank:(nr), cutoff:(e=0.1).
FEATURES             Location/Qualifiers
```



```
     source          1. .533
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone_lib="MF"
                     /tissue_type="Root"
                     /dev_stage="5-6 weeks old"
                     /note="Vector: Bluescript SK+; Site_1: EcorI; Site_2:
                     XhoI"
BASE COUNT     164 a    106 c    139 g    124 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 134;  Length 533;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttga 409
           |||||||||||||||||||||
Db     384 AAAAAAAAAAAAAATGCTTGA 404


RESULT  83
AL378418
LOCUS       AL378418      542 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBB38B03R1 MtBB Medicago truncatula cDNA clone MtBB38B03 T7, mRNA
            sequence.
ACCESSION   AL378418
VERSION     AL378418.1  GI:9678170
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Journet,E.P., Creapeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Glaninazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
            nodules
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
```



```
     source          1. .542
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBB3BB03"
                     /clone_lib="MtBB"
                     /tissue_type="symbiotic root nodules"
                     /dev_stage="harvested 4 days post inoculation with
                     Sinorhizobium meliloti"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber on
                     nitrogen-rich medium for 21 days. Three days before
                     inoculation with Sinorhizobium meliloti, the medium was
                     replaced by N-free medium. Root nodules (+ short adjacent
                     root segments) were harvested 4 days post inoculation.
                     cDNA was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zap XR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT     173 a     90 c    128 g    151 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 28;  Length 542;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db     498 ATTAATGAAAAAAAAAAAAA 518


RESULT  84
AW028928/c
LOCUS       AW028928      542 bp    mRNA            EST       15-SEP-1999
DEFINITION  wv97f08.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2542119 3'
            similar to gb:X70476 COATOMER BETA' SUBUNIT (HUMAN);, mRNA
            sequence.
ACCESSION   AW028928
VERSION     AW028928.1  GI:5887684
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```



```
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington Univeraity Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 67.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2542119"
                     /clone_lib="NCI_CGAP_Gas4"
                     /tissue_type="poorly differentiated adenocarcinoma with
                     signet ring cell features"
                     /lab_host="DH10B"
                     /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.69 kb. Life Technologies catalog #:
                     11549-011"
BASE COUNT     159 a     99 c     96 g    187 t      1 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 39;  Length 542;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db      24 ATTAATGAAAAAAAAAAAAA 4


RESULT  85
AI783160
LOCUS       AI783160      546 bp    mRNA            EST       30-JUN-1999
DEFINITION  BSBmL3SZ15E18SK Brugia malayi infective larva cDNA (SAW94WL-BmL3)
            Brugia malayi cDNA clone BSBmL3SZ15E18SK 5', mRNA sequence.
ACCESSION   AI783160
VERSION     AI783160.1  GI:5296390
KEYWORDS    EST.
SOURCE      Brugia malayi.
```

```
  ORGANISM  Brugia malayi
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Brugia.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Williams,S.A.
  TITLE     Genes expressed in infective third stage larvae of Brugia malayi
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            EST submitted by Molecular Parasitology Group, New England Biolabs,
            Inc., 32 Tozer Road, Beverly, MA. 01915, USA. Email:
            dnaseq@neb.com. SZ in the clone designation, refers to 'Subtracted
            Method Z'. The numerical designation after the SZ refers to the
            microtiter tray number (01 - 48) followed by the letter and number
            of the microtiter tray position. Method 'Z' refers to sequenced
            clones which have been selected after hybridization subtraction
            using highly redundant clones present in the gridded library (18
            ,000 mass-excised colonies gridded as a high density array on nylon
            filters). Colonies not represented in the probe sets were used as
            templates for the sequencing reactions.
            Seq primer: pBluescript SK.
FEATURES              Location/Qualifiers
     source           1. .546
                      /organism="Brugia malayi"
                      /strain="TRS Labs"
                      /db_xref="taxon:6279"
                      /clone="BSBmL3SZ15E18SK"
                      /clone_lib="Brugia malayi infective larva cDNA
                      (SAM94WL-BmL3)"
                      /lab_host="XL1-Blue MRF'"
                      /note="Vector: lambda UniZap XR; Site_1: EcoR I; Site_2:
                      Xho I; Lymphatic filarial nematode parasite of humans.
                      mRNA was prepared from third stage infective larvae of
                      Brugia malayi isolated from mosquitoes 10 days after
                      infection and converted to double stranded cDNA using
                      reverse transcriptase and oligo(dT) followed by RNase H
                      and DNApol I. The library had 1.6 x 10E6 independent
                      recombinants and average insert size was 900 base pairs.
                      The library was constructed by Wenhong-Lu. The library is
                      available from Dr. S.A. Williams, email genome@smith.edu."
BASE COUNT      191 a    102 c     70 g    182 t       1 others
ORIGIN

  Query Match             0.4%;  Score 21;  DB 25;  Length 546;
```



```
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaaatgcttga 409
           |||||||||||||||||||||
Db     229 AAAAAAAAAAAAAATGCTTGA 249


RESULT  86
AV549932
LOCUS       AV549932       551 bp    mRNA            EST       06-SEP-2000
DEFINITION  AV549932 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZ106a03R 5', mRNA sequence.
ACCESSION   AV549932
VERSION     AV549932.1  GI:8721345
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicota; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES              Location/Qualifiers
     source           1. .551
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /clone="RZ106a03R"
                      /clone_lib="Arabidopsis thaliana roots Columbia"
                      /tissue_type="roots"
                      /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                      XhoI"
BASE COUNT      155 a    117 c    133 g    146 t
ORIGIN

  Query Match             0.4%;  Score 21;  DB 37;  Length 551;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy    1800 aactgagattaaagatttcca 1820
           ||||||||||||||||||||
Db     105 AACTGAGATTAAAGATTTCCA 125


RESULT  87
B96756
LOCUS       B96756         574 bp    DNA             GSS       31-MAR-1998
DEFINITION  T32F9TR TAMU Arabidopsis thaliana genomic clone T32F9, DNA
            sequence.
ACCESSION   B96756
VERSION     B96756.1  GI:2998002
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal.Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T32F9TF
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 574.
FEATURES              Location/Qualifiers
     source           1. .574
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /clone="T32F9"
                      /clone_lib="TAMU"
                      /sex="hermaphrodite"
                      /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                      ; Produced by Rod Wing"
BASE COUNT      164 a    126 c    135 g    149 t
ORIGIN

  Query Match             0.4%;  Score 21;  DB 184;  Length 574;
```



```
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1800 aactgagattaaagatttcca 1820
           ||||||||||||||||||||
Db     141 AACTGAGATTAAAGATTTCCA 161


RESULT  88
AQ346762/c
LOCUS       AQ346762       587 bp    DNA             GSS       07-MAY-1999
DEFINITION  RPCI11-124E13.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-124E13,
            DNA sequence.
ACCESSION   AQ346762
VERSION     AQ346762.1  GI:4171658
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-124E13.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .587
                      /organism="Homo sapiens"
                      /db_xref="GDB:7547340"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-124E13"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
```

```
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       119 a      88 c      92 g     288 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 156;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     378 aatgattaatgaaaaaaaaaa 398
           ||||||||||||||||||||
Db     365 AATGATTAATGAAAAAAAAAA 345


RESULT  89
AZ235445/c
LOCUS       AZ235445      598 bp    DNA             GSS       14-JUN-2000
DEFINITION  RPCI-23-70E23.TV RPCI-23 Mus musculus genomic clone RPCI-23-70E23,
            DNA sequence.
ACCESSION   AZ235445
VERSION     AZ235445.1  GI:8543491
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-70E23.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 70  row: E  column: 23
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
```



```
     source          1. .598
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-70E23"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).
BASE COUNT       130 a      106 c      102 g      260 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 177;  Length 598;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1114 caagaagcagaaaaaaatgga 1134
           ||||||||||||||||||||
Db     212 CAAGAAGCAGAAAAAATGGA 192


RESULT  90
AV386003
LOCUS       AV386003      606 bp    mRNA            EST       27-OCT-1999
DEFINITION  AV386003 Halocynthia roretzi Pertilized egg Halocynthia roretzi
            cDNA clone 007M09_3 3', mRNA sequence.
ACCESSION   AV386003
VERSION     AV386003.1  GI:6131060
KEYWORDS    EST.
SOURCE      Halocynthia roretzi.
  ORGANISM  Halocynthia roretzi
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Stolidobranchia; Pyuridae; Halocynthia.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Makabe,K.W.
  TITLE     Halocynthia roretzi EST
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Kazuhiro W. Makabe
            Department of Zoology, Graduate School of Science
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4095
            Fax: 81-75-705-1113
            Email: kwmakabe@ascidian.zool.kyoto-u.ac.jp.
```



```
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Halocynthia roretzi"
                     /db_xref="taxon:7729"
                     /clone="007M09_3"
                     /clone_lib="Halocynthia roretzi Fertilized egg"
                     /dev_stage="Fertilized egg"
BASE COUNT       222 a      67 c      91 g     226 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 36;  Length 606;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     579 ATTAATGAAAAAAAAAAAAA 599


RESULT  91
AV699504
LOCUS       AV699504      614 bp    mRNA            EST       25-SEP-2000
DEFINITION  AV699504 GKB Homo sapiens cDNA clone GKBAIA03 3', mRNA sequence.
ACCESSION   AV699504
VERSION     AV699504.1  GI:10301475
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.
            , Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
            Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
            ,Y., Gu,Y., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA GK- clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .614
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GKBAIA03"
```



```
                     /clone_lib="GKB"
                     /tissue_type="hepatocellular carcinoma"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       131 a      146 c      119 g      218 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 38;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     593 ATTAATGAAAAAAAAAAAAA 613


RESULT  92
AA920399/c
LOCUS       AA920399      657 bp    mRNA            EST       20-APR-1998
DEFINITION  vx95g12.r1 Stratagene mouse macrophage (#937306) Mus musculus cDNA
            clone IMAGE:1293766 5' similar to gb:X16670 Mouse RNA for type IIB
            intracisternal A-particle (MOUSE);, mRNA sequence.
ACCESSION   AA920399
VERSION     AA920399.1  GI:3067178
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:674814
            Seq primer: -28m13 rev1 ET from Amersham
```

```
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1293766"
                     /clone_lib="Stratagene mouse macrophage (#937306)"
                     /tissue_type="macrophage"
                     /dev_stage="WEHI-3 cell line"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: blood; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. WEHI-3 cell line. Average insert size: 1.5 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       205 a    141 c    132 g    179 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 13;  Length 657;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5754 ttttaaaaggctttttaaaac 5774
          ||||||||||||||||||||
Db    608 TTTTAAAAGGCTTTTTAAAAC 588


RESULT  93
CNS03M8P/c
LOCUS       CNS03M8P      669 bp    DNA             GSS       17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            037I10 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL250450
VERSION     AL250450.1  GI:7971462
KEYWORDS    GSS: genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
            Holacanthopterygii; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,P., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
```



```
REFERENCE   2  (bases 1 to 669)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 669)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="037I10"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG037BE05LP1-end : T7"
BASE COUNT       199 a    144 c    153 g    142 t     31 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 192;  Length 669;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3891 tacagcatcttattactgctt 3911
          |||||||||||||||||||||
Db     63 TACAGCATCTTATTACTGCTT 43


RESULT  94
AQ491144/c
LOCUS       AQ491144      670 bp    DNA             GSS       24-APR-1999
DEFINITION  RPCI-11-265I18.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-265I18
            , DNA sequence.
ACCESSION   AQ491144
VERSION     AQ491144.1  GI:4677018
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
```



```
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-265I18.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .670
                     /organism="Homo sapiens"
                     /db_xref="GDB:7601585"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-265I18"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       156 a    168 c    137 g    209 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 160;  Length 670;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgcttga 409
          |||||||||||||||||||||
Db    319 AAAAAAAAAAAAAATGCTTGA 299


RESULT  95
AZ382945
LOCUS       AZ382945      692 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0140J23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0140J23 F, DNA sequence.
ACCESSION   AZ382945
VERSION     AZ382945.1  GI:10496645
KEYWORDS    GSS.
SOURCE      house mouse.
```



```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0140  row: J  column: 23
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 692.
FEATURES             Location/Qualifiers
     source          1. .692
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0140J23"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
```

and selected for ampicillin resistance."
BASE COUNT 270 a 101 c 94 g 227 t
ORIGIN

Query Match 0.4%; Score 21; DB 180; Length 692;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 297 atttttgttaattgtgggtaa 317
||||||||||||||||||||
Db 182 ATTTTTGTTAATTGTGGGTAA 202

RESULT 96
AQ110183
LOCUS AQ110183 712 bp DNA GSS 29-AUG-1998
DEFINITION CIT-HSP-2376C21.TR CIT-HSP Homo sapiens genomic clone 2376C21, DNA sequence.
ACCESSION AQ110183
VERSION AQ110183.1 GI:3486873
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 712)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2376C21.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2376C21"

/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII"
BASE COUNT 261 a 165 c 102 g 184 t
ORIGIN

Query Match 0.4%; Score 21; DB 149; Length 712;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2207 ataaactaaccactaaaggat 2227
||||||||||||||||||||
Db 205 ATAAACTAACCACTAAAGGAT 225

RESULT 97
AZ183754
LOCUS AZ183754 722 bp DNA GSS 30-AUG-2000
DEFINITION SP_1001_B2_F06_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1001 Col=12 Row=L, DNA sequence.
ACCESSION AZ183754
VERSION AZ183754.1 GI:8356129
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 722)
AUTHORS Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and Hood,L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
MEDLINE 20402566
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 1001 row: L column: 12
Seq primer: SP6
Class: BAC ends

High quality sequence stop: 722.
FEATURES Location/Qualifiers
source 1. .722
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=1001 Col=12 Row=L"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"
BASE COUNT 165 a 121 c 161 g 275 t
ORIGIN

Query Match 0.4%; Score 21; DB 176; Length 722;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 tattgatttgttttccatttt 89
||||||||||||||||||||
Db 252 TATTGATTTGTTTTCCATTTT 272

RESULT 98
CNS0252V
LOCUS CNS0252V 742 bp DNA GSS 12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 236I07 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL181552
VERSION AL181552.1 GI:7819629
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha; Holacanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 742)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 742)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 742)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .742
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="236I07"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG236AE04LP1-end : T7"
BASE COUNT 307 a 124 c 120 g 157 t 34 others
ORIGIN

Query Match 0.4%; Score 21; DB 191; Length 742;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 attaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 507 ATTAATGAAAAAAAAAAAAAA 527

RESULT 99
AZ195068/c
LOCUS AZ195068 757 bp DNA GSS 30-AUG-2000
DEFINITION SP_1029_A2_E07_T7A Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1029 Col=14 Row=I, DNA sequence.
ACCESSION AZ195068
VERSION AZ195068.1 GI:8378247
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 757)
AUTHORS Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and Hood,L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources

```
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1029  row: I  column: 14
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 757.
FEATURES              Location/Qualifiers
     source           1. .757
                      /organism="Strongylocentrotus purpuratus"
                      /db_xref="taxon:7668"
                      /clone="Plate=1029 Col=14 Row=I"
                      /clone_lib="Strongylocentrotus purpuratus, purple sea
                      urchin, sperm genomic BAC library"
                      /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                      DH10B"
BASE COUNT      192 a      93 c     136 g     333 t       3 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 176;  Length 757;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgcttga 409
          ||||||||||||||||||||
Db    403 AAAAAAAAAAAAAATGCTTGA 383


RESULT 100
BE269463/c
LOCUS       BE269463      781 bp    mRNA            EST       13-JUL-2000
DEFINITION  601184638F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3542528 5',
            mRNA sequence.
ACCESSION   BE269463
VERSION     BE269463.1  GI:9143084
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
```



```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM234  row: k  column: 09
            High quality sequence start: 3
            High quality sequence stop: 692.
FEATURES              Location/Qualifiers
     source           1. .781
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3542528"
                      /clone_lib="NIH_MGC_8"
                      /tissue_type="Burkitt lymphoma"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Size-selected >500bp for average
                      insert size 1.8kb.  Library constructed by Ling Hong in
                      the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using  ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      162 a    184 c    265 g    170 t
ORIGIN


  Query Match             0.4%;  Score 21;  DB 106;  Length 781;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    382 attaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     25 ATTAATGAAAAAAAAAAAAAA 5


RESULT 101
AZ046289/c
LOCUS       AZ046289      791 bp    DNA             GSS       16-MAR-2000
DEFINITION  nbeb0091J15r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone OSJNBb0091J15r, DNA sequence.
ACCESSION   AZ046289
VERSION     AZ046289.1  GI:7250058
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
```



```
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GCTATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 29
            High quality sequence stop: 262.
FEATURES              Location/Qualifiers
     source           1. .791
                      /organism="Oryza sativa"
                      /strain="Japonica"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:4530"
                      /clone="OSJNBb0091J15r"
                      /clone_lib="CUGI Rice BAC Library (EcoRI)"
                      /tissue_type="Leaf"
                      /lab_host="E. coli DH10B"
                      /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                      Rice is the most important food crop in the world. Half of
                      the world population, especially those inhabiting highly
                      populated areas of the humid tropics and subtropics, rely
                      on rice as their primary source of carbohydrate.
                      Monocotyledonous rice is a diploid plant (2n=24) with a
                      haploid genome equivalent of 431 Mbp (Arumuganathan and
                      Earle, 1991). The relatively small genome of rice, three
                      times larger than that of Arabidopsis, makes it suitable
                      for genomic studies. In order to facilitate positional
                      cloning, physical mapping and genome sequencing of rice,
                      we have constructed a BAC library from Oryza sativa,
                      Nipponbare variety using EcoRI as the cloning enzyme. The
                      library contains 55,296 clones with an average insert size
                      of 121 Kb providing approximatley 15 haploid genome
                      equivalents. The deep coverage allows the isolation a
                      particular sequence with a probability of 99.9 %. Three
                      high density filters, each containing 18,432 clones
                      (doubly spotted), represent the whole library for colony
                      screening and can be requested from the Clemson University
                      BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT       59 a    179 c    131 g    307 t    115 others
ORIGIN
```



```
  Query Match             0.4%;  Score 21;  DB 174;  Length 791;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1773 gagcgagaaagagaaggaaaa 1793
          ||||||||||||||||||||
Db    752 GAGCGAGAAAGAGAAGGAAAA 732


RESULT 102
CNS01ZTJ
LOCUS       CNS01ZTJ      955 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            221K23 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL174736
VERSION     AL174736.1  GI:7812793
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
            Holacanthopterygii; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 955)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 955)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 955)
  AUTHORS   Genoscope
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES              Location/Qualifiers
     source           1. .955
                      /organism="Tetraodon nigroviridis"
```

```
                     /db_xref="taxon:99883"
                     /clone="221K23"
                     /clone_lib="C"
                     /note="Genoscope sequence ID : C0AG221AF12LP1-end : T7"
BASE COUNT       289 a    190 c    215 g    256 t      5 others
ORIGIN


  Query Match             0.4%;  Score 21;  DB 191;  Length 955;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      52 tctttgaaaatctgtcctatt 72
           |||||||||||||||||||||
Db     603 TCTTTGAAAATCTGTCCTATT 623


RESULT 103
AI963429/c
LOCUS       AI963429       65 bp    mRNA            EST       09-MAR-2000
DEFINITION  wt63b05.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2512113 3',
            mRNA sequence.
ACCESSION   AI963429
VERSION     AI963429.1  GI:5756142
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 2123    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 55.
FEATURES             Location/Qualifiers
     source          1. .65
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2512113"
                     /clone_lib="NCI_CGAP_Pan1"
```



```
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT       10 a      4 c     16 g     35 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 27;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      37 TTAATGAAAAAAAAAAAAAA 18


RESULT 104
AA110380
LOCUS       AA110380       76 bp    mRNA            EST       02-OCT-1997
DEFINITION  mp43c03.r1 Barstead MPLRB1 Mus musculus cDNA clone IMAGE:571972 5'
            similar to TR:G473969 G473969 ONE OF THE MEMBERS OF SODIUM-GLUCOSE
            COTRANSPORTER FAMILY. :, mRNA sequence.
ACCESSION   AA110380
VERSION     AA110380.1  GI:1662124
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 76)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI: 346620
            Possible reversed clone: similarity on wrong strand
```



```
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .76
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:571972"
                     /clone_lib="Barstead MPLRB1"
                     /sex="mixed"
                     /tissue_type="Kidney"
                     /dev_stage="6 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [CATGGATTCGGTACC], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead."
BASE COUNT       20 a      15 c     21 g     20 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 2;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4227 acacagagtgagaaccaaac 4246
           ||||||||||||||||||||
Db     52 ACACAGAGTGAGAACCAAAC 71


RESULT 105
AI913476/c
LOCUS       AI913476       82 bp    mRNA            EST       16-DEC-1999
DEFINITION  tz77h02.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2294643 3',
            mRNA sequence.
ACCESSION   AI913476
VERSION     AI913476.1  GI:5633331
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 82)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```



```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 2287    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 63.
FEATURES             Location/Qualifiers
     source          1. .82
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2294643"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT       12 a      4 c     17 g     49 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 26;  Length 82;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      51 TTAATGAAAAAAAAAAAAAA 32


RESULT 106
AA399932
LOCUS       AA399932      109 bp    mRNA            EST       28-APR-1997
DEFINITION  vd67f07.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:805669 5', mRNA sequence.
ACCESSION   AA399932
VERSION     AA399932.1  GI:2053388
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 109)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
```

```
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:482013
            Putative full length read
            vector to vector length is 121
            High quality sequence stop: 85.
FEATURES              Location/Qualifiers
     source           1. .109
                      /organism="Mus musculus"
                      /strain="B6D2 F1/J"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:805669"
                      /clone_lib="Knowles Solter mouse blastocyst B1"
                      /tissue_type="blastocyst"
                      /dev_stage="embryo (pre-implantation)"
                      /lab_host="DH10B"
                      /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally from mRNA prepared
                      from 800 blastocysts.  Primer: SalI(dT):
                      5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'.  cDNAs  were
                      cloned into the NotI/SalI sites of a pSPORT vector (Life
                      Technologies).  Two different size selections:  B1 (larger
                      inserts) and B3."
BASE COUNT       58 a      11 c      22 g      18 t
ORIGIN


  Query Match           0.3%;  Score 20;  DB 7;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      86 TTAATGAAAAAAAAAAAAAA 105


RESULT 107
AW333457/c
LOCUS       AW333457      120 bp    mRNA            EST       31-JAN-2000
```



```
DEFINITION  S22A9 AGS-1 Pneumocystis carinii f. sp. carinii cDNA 3', mRNA
            sequence.
ACCESSION   AW333457
VERSION     AW333457.1  GI:6829814
KEYWORDS    EST.
SOURCE      Pneumocystis carinii f. sp. carinii.
  ORGANISM  Pneumocystis carinii f. sp. carinii
            Eukaryota; Fungi; Fungi incertae sedis; Pneumocystidaceae;
            Pneumocystis.
REFERENCE   1  (bases 1 to 120)
  AUTHORS   Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman
            ,J.C., Kovacs,J. and Cushion,M.
  TITLE     Expressed sequence tags from Pneumocystis carinii
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Staben C
            School of Biological Sciences
            University of Kentucky
            101 Morgan Building, University of Kentucky, Lexington, KY
            40506-0225, USA
            Tel: 606 257 2161
            Fax: 606 257 1717
            Email: staben@pop.uky.edu.
FEATURES              Location/Qualifiers
     source           1. .120
                      /organism="Pneumocystis carinii f. sp. carinii"
                      /db_xref="taxon:38081"
                      /clone_lib="AGS-1"
                      /lab_host="E. coli"
                      /note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
                      P. carinii organisms (3x10e9) from a single rat (99-1-6,
                      sacrificed on 3/17/99)  at Cincinnati VA facilities.
                      Trizol extracted RNA.  Oligo dT priming, standard
                      conditions described by  vendor, Stratagene. Further
                      details see www.uky.edu/Project/Pneumocystis/"
BASE COUNT       54 a       8 c      11 g      47 t
ORIGIN


  Query Match           0.3%;  Score 20;  DB 89;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      26 TTAATGAAAAAAAAAAAAAA 7


RESULT 108
AI537111/c
LOCUS       AI537111      122 bp    mRNA            EST       12-MAY-1999
DEFINITION  to15h03.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2179157 3',
```



```
            mRNA sequence.
ACCESSION   AI537111
VERSION     AI537111.1  GI:4451246
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 122)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 4083  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 84
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .122
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2179157"
                      /clone_lib="NCI_CGAP_Ut2"
                      /tissue_type="moderately-differentiated endometrial
                      adenocarcinoma, 3 pooled tumors"
                      /lab_host="DH10B"
                      /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.85 kb. Life Technologies catalog #:
                      11539-012"
BASE COUNT       42 a       9 c      13 g      58 t
ORIGIN


  Query Match           0.3%;  Score 20;  DB 21;  Length 122;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
```



```
Db      36 TTAATGAAAAAAAAAAAAAA 17


RESULT 109
BE104648/c
LOCUS       BE104648      128 bp    mRNA            EST       13-JUN-2000
DEFINITION  UI-R-BX0-arr-h-05-0-UI.s1 UI-R-BX0 Rattus norvegicus cDNA clone
            UI-R-BX0-arr-h-05-0-UI 3', mRNA sequence.
ACCESSION   BE104648
VERSION     BE104648.1  GI:8496748
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 128)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized embryo at 13 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .128
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-BX0-arr-h-05-0-UI"
                      /clone_lib="UI-R-BX0"
                      /dev_stage="embryonic 13 dpc"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BX0
                      library is derived from 13 dpc whole embryo tissue. For a
```

detailed description of the library from which this clone
was derived, please visit our web site at
ratest.eng.uiowa.edu.
TAG_LIB=UI-R-BX0
TAG_TISSUE=embryo at 13 dpc
TAG_SEQ=AATCC"
BASE COUNT 36 a 24 c 21 g 47 t
ORIGIN

Query Match 0.3%; Score 20; DB 104; Length 128;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAAA 4

RESULT 110
AI540717/c
LOCUS AI540717 130 bp mRNA EST 14-APR-1999
DEFINITION tn72f01.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2175097 3' similar to gb:X16869 ELONGATION FACTOR 1-ALPHA 1 (HUMAN);, mRNA sequence.
ACCESSION AI540717
VERSION AI540717.1 GI:4458090
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 130)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Life Technologies catalog #: 11547-015
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 355 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
source 1. .130
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="IMAGE:2175097"
/clone_lib="NCI_CGAP_Lym12"
/tissue_type="lymphoma, follicular mixed small and large cell"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.25 kb. Life Technologies catalog #: 11547-015"
BASE COUNT 38 a 21 c 13 g 58 t
ORIGIN

Query Match 0.3%; Score 20; DB 21; Length 130;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 25 TTAATGAAAAAAAAAAAAAA 6

RESULT 111
AW576296/c
LOCUS AW576296 134 bp mRNA EST 15-MAR-2000
DEFINITION UI-HF-BN0-aki-c-02-0-UI.s1 NIH_MGC_50 Homo sapiens cDNA clone IMAGE:3076995 3', mRNA sequence.
ACCESSION AW576296
VERSION AW576296.1 GI:7247835
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 134)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
cDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .134
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3076995"
/clone_lib="NIH_MGC_50"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (3.5-4.4kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D. "
BASE COUNT 43 a 16 c 21 g 54 t
ORIGIN

Query Match 0.3%; Score 20; DB 92; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAAA 4

RESULT 112
AW444585/c
LOCUS AW444585 148 bp mRNA EST 17-FEB-2000
DEFINITION UI-H-BI3-ajv-f-09-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone IMAGE:2733089 3', mRNA sequence.
ACCESSION AW444585
VERSION AW444585.1 GI:6986347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 148)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .148
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2733089"
/clone_lib="NCI_CGAP_Sub5"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub5 is a subtracted library derived from NCI_CGAP_Sub4. The NCI_CGAP_Sub5 library had 3 million recombinants. A single-stranded DNA preparation of NCI_CGAP_Sub4 was used as a tracer in a subtractive hybridization with a driver comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs 1322376-1323911, 1456008-1456775,1500552-1502855); NCI_CGAP_Kid5 pool 1 LLAM 3338-3342,3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831, 1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759 ,1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255 ,1144584-1145351). (10% of the driver population), plus a pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE CloneIDs 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE CloneIDs 2710536-2712455) (10% of the driver population), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE CloneIDs 2712456-2723591) (10% of the driver population), plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE CloneIDs 2723592-2728969) (70% of the driver population). Subtraction was performed as previously described [Bonaldo , Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_LIB=NCI_CGAP_Lei2
TAG_TISSUE=leiomyosarcoma
TAG_SEQ=AATCG"
BASE COUNT 47 a 25 c 30 g 46 t
ORIGIN

```
  Query Match          0.3%;  Score 20;  DB 90;  Length 148;
  Best Local Similarity  100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     23 TTAATGAAAAAAAAAAAAA 4


RESULT 113
AW138945/c
LOCUS       AW138945      160 bp    mRNA            EST       29-OCT-1999
DEFINITION  UI-H-BI1-ady-a-07-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
            IMAGE:2718252 3', mRNA sequence.
ACCESSION   AW138945
VERSION     AW138945.1  GI:6143263
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html The following repetitive
            elements were found in this cDNA sequence: 23-159, >MLT1D#LTR/MaLR
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..160
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2718252"
                     /clone_lib="NCI_CGAP_Sub3"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NCI_CGAP_Sub3 library is a subtracted library derived from
```



```
                     the  NCI_CGAP_Sub1 library, which is a subtracted library
                     derived from BI.  BI constitutes a mixture of 21
                     normalized or subtracted NCI_CGAP  libraries: NCI_CGAP_Co4
                     , NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
                     NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
                     NCI_CGAP_Kid3,  NCI_CGAP_Kid11, NCI_CGAP_Lym2,
                     NCI_CGAP_Br2, NCI_CGAP_Co8,  NCI_CGAP_CLL1, NCI_CGAP_Lei2,
                     NCI_CGAP_Brn23, NCI_CGAP_Lu5,  NCI_CGAP_Lu24,
                     NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
                     NCI_CGAP_Brn25.  These 21 libraries were pooled and a
                     single-stranded DNA  preparation of the resulting mixture
                     was used as a tracer in a subtractive  hybridization with
                     a driver whose composition is detailed below:
                     NCI_CGAP_Kid3 pool 1     LLAM 3334-3337, 3682-3683,
                     3798-3803 (IMAGE CloneIDs 1322376-1323911, 1456008-1456775
                     , 1500552-1502855); NCI_CGAP_Kid5 pool 1     LLAM 3338-3342
                     , 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831,
                     1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1
                     LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,
                     1520904-1522439); NCI_CGAP_GC4 pool 1.     LLAM 3164-3167,
                     3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,
                     1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1
                     LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs
                     985608-986759, 1101192-1101959, 1217928-1220615);
                     NCI_CGAP_Co10 pool 1     LLAM 2644-2653, 2871-2872 (IMAGE
                     CloneIDs 1057416-1061255, 1144584-1145351). Subtraction
                     was performed as previously described [Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To  Facilitate Gene Discovery. Genome Research
                     6, 791-806.
                     TAG_LIB=NCI_CGAP_Lu5
                     TAG_TISSUE=lung
                     TAG_SEQ=CAAC"
BASE COUNT       35 a      34 c      35 g      56 t
ORIGIN


  Query Match          0.3%;  Score 20;  DB 40;  Length 160;
  Best Local Similarity  100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     23 TTAATGAAAAAAAAAAAAA 4


RESULT 114
BB170597/c
LOCUS       BB170597      167 bp    mRNA            EST       29-JUN-2000
DEFINITION  BB170597 RIKEN full-length enriched, adult male hypothalamus Mus
            musculus cDNA clone A230020A15 3', mRNA sequence.
```



```
ACCESSION   BB170597
VERSION     BB170597.1  GI:8829680
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 167)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..167
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
```



```
                     /clone="A230020A15"
                     /clone_lib="RIKEN full-length enriched, adult male
                     hypothalamus"
                     /sex="male"
                     /tissue_type="hypothalamus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. "
BASE COUNT       64 a      15 c      23 g      65 t
ORIGIN


  Query Match          0.3%;  Score 20;  DB 102;  Length 167;
  Best Local Similarity  100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgcttg 408
          ||||||||||||||||||||
Db     89 AAAAAAAAAAAAAATGCTTG 70


RESULT 115
AV378192/c
LOCUS       AV378192      177 bp    mRNA            EST       15-NOV-1999
DEFINITION  AV378192 RIKEN full-length enriched, adult male cecum Mus musculus
            cDNA clone 9130411I07 3', mRNA sequence.
ACCESSION   AV378192
VERSION     AV378192.1  GI:6427201
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 177)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
```

```
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .177
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="9130411I07"
                      /clone_lib="RIKEN full-length enriched, adult male cecum"
                      /sex="male"
                      /tissue_type="cecum"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
```



```
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 10.0 and subtraction to Rot = 185.0. Second
                      strand cDNA was prepared with the primer adapter of
                      sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                      3']. cDNA was cloned into the XhoI and BamHI sites.
                      Vector: a modified pBluescript KS(+) after bulk excision
                      from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                      BamHI"
BASE COUNT       43 a      30 c      36 g      68 t      . .
ORIGIN


  Query Match             0.3%;  Score 20;  DB 36;  Length 177;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     140 TTAATGAAAAAAAAAAAAAA 121


RESULT 116
AI688140/c
LOCUS       AI688140      190 bp    mRNA            EST       07-MAR-2000
DEFINITION  wa83g10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2302818 3' similar to contains L1.t1 L1 repetitive element ;,
            mRNA sequence.
ACCESSION   AI688140
VERSION     AI688140.1  GI:4899434
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 686   Std Error: 0.00
```



```
            Seq primer: -40UP from Gibco
            High quality sequence stop: 188.
FEATURES              Location/Qualifiers
     source           1. .190
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2302818"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo. "
BASE COUNT       55 a      32 c      40 g      63 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 23;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      31 TTAATGAAAAAAAAAAAAAA 12


RESULT 117
BE808167
LOCUS       BE808167      192 bp    mRNA            EST       20-SEP-2000
DEFINITION  213308 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE808167
VERSION     BE808167.1  GI:10239279
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 192)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
```



```
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 67  row: D  column: 13
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .192
                      /organism="Bos taurus"
                      /db_xref="taxon:9913".
                      /clone_lib="MARC 2BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from testis, thymus,
                      semitendonosus muscle, longissimus muscle, pancreas,
                      adrenal, and endometrium."
BASE COUNT       50 a      45 c      57 g      40 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 136;  Length 192;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaatg 404
           ||||||||||||||||||||
Db     169 AATGAAAAAAAAAAAAAATG 188


RESULT 118
BE946267/c
LOCUS       BE946267      194 bp    mRNA            EST       03-OCT-2000
DEFINITION  UI-M-BH3-awj-h-03-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-awj-h-03-0-UI 3', mRNA sequence.
ACCESSION   BE946267
VERSION     BE946267.1  GI:10524026
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 194)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            pineal glands tissue cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: Researchers may obtain BMAP cDNA clones from
            RESEARCH GENETICS. It should be noted that Bento Soares is
            generating a small number of additional specialized non-redundant
            arrays of BMAP cDNAs whose availability will be considered under
            appropriate and limited collaborative arrangements The following
            repetitive elements were found in this cDNA sequence: 1-29,
            >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .194
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-BH3-awj-h-03-0-UI"
                      /clone_lib="NIH_BMAP_M_S4"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_M_S4 library is a subtracted library of a series,
                      ultimately  derived from a mixture of individually tagged
                      normalized libraries from  ten regions of the mouse brain
                      (cerebellum, brain stems, olfactory  bulbs, hypothalamus,
                      cortex, amygdala, basal ganglia, pineal gland,  striatum,
                      hipoccampus) after a series of subtractions to reduce the
                      representation of cDNAs from which ESTs had already been
                      generated. The following serially subtracted libraries
                      were generated in this process: NIH_BMAP_M_S4,
                      NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                      NIH_BMAP_M_S2, NIH_BMAP_M_S1.  The subtracted library
```



```
                      (NIH_BMAP_M_S4) was constructed as follows:  PCRamplified
                      cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                      NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                      was used as  a driver in a hybridization with a pool of
                      the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                      libraries in the form of  single-stranded circles.  The
                      remaining single-stranded circles  (subtracted library)
                      was purified by hydroxyapatite column chromatography,
                      converted to double-stranded circles and electroporated
                      into DH10B  bacteria (LifeTechnologies) to generate the
                      NIH_BMAP_M_S4 library.  This procedure has been previously
                      described (Bonaldo, Lennon and Soares, Genome Research
                      6:791-806, 1996)
                      TAG_LIB=NIH_BMAP_M_S4
                      TAG_TISSUE=pineal-glands
                      TAG_SEQ=CAGAC"
BASE COUNT       50 a      22 c      45 g      77 t
ORIGIN


  Query Match              0.3%;  Score 20;  DB 137;  Length 194;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      23 TTAATGAAAAAAAAAAAAAA 4


RESULT 119
AI059741/c
LOCUS       AI059741       201 bp    mRNA            EST       11-FEB-1999
DEFINITION  UI-R-C1-lk-f-11-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
            UI-R-C1-lk-f-11-0-UI 3', mRNA sequence.
ACCESSION   AI059741
VERSION     AI059741.1  GI:3333518
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
```



```
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult Spleen library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .201
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-C1-lk-f-11-0-UI"
                      /clone_lib="UI-R-C1"
                      /dev_stage="Adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
                      library is a subtracted library derived from the UI-R-C0
                      library, which is a subtracted library derived from the
                      UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
                      consisted of a mixture of  individually tagged normalized
                      libraries constructed from rat placenta,  adult lung,
                      brain, liver, kidney, heart, spleen, ovary, and muscle.
                      The UI-R-E1 library  consisted of a mixture of
                      individually tagged normalized libraries constructed from
                      8, 12 and 18-day embryo. The tag is a string of 3-5
                      nucleotides present between the Not I site and the
                      oligo-dT track which allows identification of the library
                      of origin of a clone within the mixture. The subtracted
                      library (UI-R-C1) was constructed as follows: PCR
                      amplified cDNA inserts from  UI-R-C0 clones from which 3'
                      ESTs had been derived was used as a driver in a
                      hybridization with the UI-R-C0 library in the form of
                      single-stranded circles. The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-C1
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
BASE COUNT       48 a      49 c      32 g      72 t
ORIGIN


  Query Match              0.3%;  Score 20;  DB 15;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
```



```
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      23 TTAATGAAAAAAAAAAAAAA 4


RESULT 120
AW357010
LOCUS       AW357010       201 bp    mRNA            EST       14-JUL-2000
DEFINITION  39751 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW357010
VERSION     AW357010.1  GI:6861016
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 20  row: D  column: 18
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .201
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 2BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from testis, thymus,
                      semitendonosus muscle, longissimus muscle, pancreas,
                      adrenal, and endometrium."
```

```
BASE COUNT       82 a     22 c     26 g     71 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 89;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     180 TTAATGAAAAAAAAAAAAAA 199


RESULT 121
H63232
LOCUS       H63232       202 bp    mRNA            EST       11-OCT-1995
DEFINITION  yr51a02.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:208778 3', mRNA sequence.
ACCESSION   H63232
VERSION     H63232.1  GI:1018033
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1328
            High quality sequence stops: 133
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1328   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 133.
FEATURES             Location/Qualifiers
     source          1. .202
                     /organism="Homo sapiens"
                     /db_xref="GDB:3777909"
                     /db_xref="taxon:9606"
```



```
                     /clone="IMAGE:208778"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       77 a     35 c     22 g     66 t      2 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 142;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db       7 TTAATGAAAAAAAAAAAAAA 26


RESULT 122
AI247269/c
LOCUS       AI247269      205 bp    mRNA            EST       28-JAN-1999
DEFINITION  qj60f09.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1863881 3',
            mRNA sequence.
ACCESSION   AI247269
VERSION     AI247269.1  GI:3842666
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 205)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
```



```
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 951   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 170.
FEATURES             Location/Qualifiers
     source          1. .205
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1863881"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia),  digested with Not I and cloned into the Not
                     I and Eco RI sites  of the modified pT7T3 vector.  mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo."
BASE COUNT       72 a     23 c     19 g     91 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 18;  Length 205;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      21 TTAATGAAAAAAAAAAAAAA 2


RESULT 123
C84740
LOCUS       C84740       206 bp    mRNA            EST       12-JUL-1999
DEFINITION  C84740 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSF171, mRNA sequence.
ACCESSION   C84740
VERSION     C84740.1  GI:2864887
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 206)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
```



```
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .206
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSF171"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      108 a     18 c     19 g     58 t      3 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 140;  Length 206;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     140 TTAATGAAAAAAAAAAAAAA 159


RESULT 124
AA903342/c
LOCUS       AA903342      207 bp    mRNA            EST       08-APR-1998
DEFINITION  ok45h10.s1 NCI_CGAP_Lei2 Homo sapiens cDNA clone IMAGE:1516963 3',
            mRNA sequence.
ACCESSION   AA903342
VERSION     AA903342.1  GI:3038465
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 207)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 164.
FEATURES              Location/Qualifiers
     source           1. .207
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1516963"
                      /clone_lib="NCI_CGAP_Lei2"
                      /tissue_type="leiomyosarcoma"
                      /lab_host="DH10B"
                      /note="Organ: soft tissue; Vector: pT7T3D-Pac (Pharmacia)
                      with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      1st strand cDNA was primed with a Not I - oligo(dT) primer
                      [5'-AACTGGAAGAATTCGCGGCCGCAATCGTTTTTTTTTTTTTTTTT-3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       71 a      42 c      17 g      77 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 13;  Length 207;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     20 TTAATGAAAAAAAAAAAAA 1


RESULT 125
BE995735/c
LOCUS       BE995735       210 bp    mRNA            EST      05-OCT-2000
DEFINITION  UI-M-CG0p-bib-h-09-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
            UI-M-CG0p-bib-h-09-0-UI 3', mRNA sequence.
ACCESSION   BE995735
VERSION     BE995735.1  GI:10680287
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 210)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
```



```
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            retina tissue cDNA Library Preparation: M.B. Soares Lab Clone
            distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
            GENETICS. It should be noted that Bento Soares is generating a
            small number of additional specialized non-redundant arrays of BMAP
            cDNAs whose availability will be considered under appropriate and
            limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .210
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-CG0p-bib-h-09-0-UI"
                      /clone_lib="NIH_BMAP_Ret4_S2"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_Ret4_S2 library is a subtracted library,
                      ultimately derived from mouse retina tissue libraries at
                      various stages of development. For a detailed description
                      of the library from which this clone was derived, please
                      visit our web site at brainest.eng.uiowa.edu.
                      TAG_LIB=NIH_BMAP_Ret4_S2
                      TAG_TISSUE=adult-retina
                      TAG_SEQ=GTCAGCGCGCAC"
BASE COUNT       58 a      33 c      37 g      82 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 138;  Length 210;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     24 TTAATGAAAAAAAAAAAAA 5
```



```
RESULT 126
T10769/c
LOCUS       T10769         215 bp    mRNA            EST       29-NOV-1993
DEFINITION  hbc1239 Human pancreatic islet Homo sapiens cDNA clone hbc1239
            3'end, mRNA sequence.
ACCESSION   T10769
VERSION     T10769.1  GI:390923
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Takeda,J., Yano,H., Eng,S., Zeng,Y. and Bell,G.I.
  TITLE     A molecular inventory of human pancreatic islets: sequence analysis
            of 1000 cDNA clones
  JOURNAL   Hum. Mol. Genet. 2, 1793-1798 (1993)
  MEDLINE   94108427
COMMENT     Contact: Bell GI or Takeda J
            HHMI
            Univ. of Chicago
            5841 S. Maryland Ave.,MC1028,Chicago IL 60637
            Tel: 3127029116
            Fax: 3127020271
            Email: g-bell@uchicago.edu
            Seq primer: T7 primer.
FEATURES              Location/Qualifiers
     source           1. .215
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="hbc1239"
                      /clone_lib="Human pancreatic islet"
                      /note="Vector: Lambda ZAPII; Site_1: Eco RI; Site_2: Xho
                      I; mRNA was prepared from normal adult human islets. cDNA
                      was directionally synthesized from the Xho I in the vector
                      to the EcoRI site. cDNA was size fractionated to remove
                      sequences <1000 bp in size."
BASE COUNT       75 a      26 c      22 g      91 t       1 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 145;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     24 TTAATGAAAAAAAAAAAAA 5
```



```
RESULT 127
BB077862/c
LOCUS       BB077862       221 bp    mRNA            EST       27-JUN-2000
DEFINITION  BB077862 RIKEN full-length enriched, adult male diencephalon Mus
            musculus cDNA clone 9330146M21 3', mRNA sequence.
ACCESSION   BB077862
VERSION     BB077862.1  GI:8642922
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 221)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
```

```
            Please visit our web site (http://genome.rtc.riken.go.jp) for
           further details.
FEATURES             Location/Qualifiers
     source          1..221
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9330146M21"
                     /clone_lib="RIKEN full-length enriched, adult male
                     diencephalon"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT       75 a      29 c      28 g      88 t       1 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 99;  Length 221;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db     143 AAAAAAAAAAAAAATGCTTG 124


RESULT 128
AV346494/c
LOCUS       AV346494      225 bp    mRNA            EST       12-NOV-1999
DEFINITION  AV346494 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430580D06 3', mRNA sequence.
ACCESSION   AV346494
```



```
VERSION     AV346494.1  GI:6387553
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 225)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..225
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6430580D06"
                     /clone_lib="RIKEN full-length enriched, adult male
```



```
                     olfactory bulb"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT       82 a      27 c      29 g      87 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 35;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db     147 AAAAAAAAAAAAAATGCTTG 128


RESULT 129
AI544518
LOCUS       AI544518      227 bp    mRNA            EST       22-MAR-1999
DEFINITION  fb76b11.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA
            sequence.
ACCESSION   AI544518
VERSION     AI544518.1  GI:4461891
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 227)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
```



```
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T7 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..227
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']);
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT       83 a      45 c      35 g      64 t
ORIGIN
```

```
  Query Match             0.3%;  Score 20;  DB 21;  Length 227;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     29 TTAATGAAAAAAAAAAAAAA 48


RESULT 130
AW050394/c
LOCUS       AW050394       227 bp    mRNA            EST       20-SEP-1999
DEFINITION  wz11f05.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2557761 3'
            similar to gb:X16869 ELONGATION FACTOR 1-ALPHA 1 (HUMAN);, mRNA
            sequence.
ACCESSION   AW050394
VERSION     AW050394.1  GI:5912709
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 227)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 153.
FEATURES             Location/Qualifiers
     source          1. .227
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2557761"
                     /clone_lib="NCI_CGAP_Ut4"
                     /tissue_type="serous papillary carcinoma, high grade, 2
                     pooled tumors"
                     /lab_host="DH10B"
```



```
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.48 kb. Life Technologies catalog #:
                     11542-016"
BASE COUNT       63 a     41 c     24 g     98 t      1 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 39;  Length 227;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     44 TTAATGAAAAAAAAAAAAAA 25


RESULT 131
AV354500/c
LOCUS       AV354500       229 bp    mRNA            EST       12-NOV-1999
DEFINITION  AV354500 RIKEN full-length enriched, 2 days neonate sympathetic
            ganglion Mus musculus cDNA clone 7120438J23 3', mRNA sequence.
ACCESSION   AV354500
VERSION     AV354500.1  GI:6395557
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 229)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
```



```
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .229
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="7120438J23"
                     /clone_lib="RIKEN full-length enriched, 2 days neonate
                     sympathetic ganglion"
                     /sex="mixed"
                     /tissue_type="sympathetic ganglion"
                     /dev_stage="2 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGAGGATCCAAGACCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I.
                     Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
                     -RNA was provided by Akira Nakagawara, Div. of
                     Biochemistry, Chiba Cancer Center Research Institute,
                     666-2 Nitona, Chuoh-ku, Chiba, 260-8717 Japan, whose
                     assistance we gratefully acknowledge."
BASE COUNT       64 a     22 c     39 g     104 t
ORIGIN
```



```
  Query Match             0.3%;  Score 20;  DB 35;  Length 229;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     67 TTAATGAAAAAAAAAAAAAA 48


RESULT 132
BB077631/c
LOCUS       BB077631       230 bp    mRNA            EST       27-JUN-2000
DEFINITION  BB077631 RIKEN full-length enriched, adult male diencephalon Mus
            musculus cDNA clone 9330142D24 3', mRNA sequence.
ACCESSION   BB077631
VERSION     BB077631.1  GI:8642691
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 230)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
```

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .230
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9330142D24"
                     /clone_lib="RIKEN full-length enriched, adult male
                     diencephalon"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT       82 a      31 c      29 g      88 t
ORIGIN
```

```
  Query Match             0.3%;  Score 20;  DB 99;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
```



```
Db     152 AAAAAAAAAAAAAATGCTTG 133


RESULT 133
BE991125/c
LOCUS       BE991125       233 bp    mRNA            EST       05-OCT-2000
DEFINITION  UI-M-BZ1-bdr-b-04-0-UI.s1 NIH_BMAP_MHI2_S1 Mus musculus cDNA clone
            UI-M-BZ1-bdr-b-04-0-UI 3', mRNA sequence.
ACCESSION   BE991125
VERSION     BE991125.1  GI:10670281
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bento Soares
            is generating a small number of additional specialized
            non-redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            The following repetitive elements were found in this cDNA sequence:
            181-210, >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BZ1-bdr-b-04-0-UI"
                     /clone_lib="NIH_BMAP_MHI2_S1"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MHI2_S1 library is a subtracted library derived
```



```
                     from  NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
                     from mouse hippocampus tissue. For a detailed description
                     of the library from which this clone was  derived, please
                     visit our web site at brainest.eng.uiowa.edu.
                     TAG_SEQ=None found"
BASE COUNT       31 a      77 c      48 g      76 t       1 others
ORIGIN
```

```
  Query Match             0.3%;  Score 20;  DB 138;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     208 TTAATGAAAAAAAAAAAAAA 189


RESULT 134
BE991210/c
LOCUS       BE991210       233 bp    mRNA            EST       05-OCT-2000
DEFINITION  UI-M-BZ1-bds-c-09-0-UI.s1 NIH_BMAP_MHI2_S1 Mus musculus cDNA clone
            UI-M-BZ1-bds-c-09-0-UI 3', mRNA sequence.
ACCESSION   BE991210
VERSION     BE991210.1  GI:10670483
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bento Soares
            is generating a small number of additional specialized
            non-redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            The following repetitive elements were found in this cDNA sequence:
```



```
            181-210, >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BZ1-bds-c-09-0-UI"
                     /clone_lib="NIH_BMAP_MHI2_S1"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MHI2_S1 library is a subtracted library derived
                     from  NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
                     from mouse hippocampus tissue. For a detailed description
                     of the library from which this clone was  derived, please
                     visit our web site at brainest.eng.uiowa.edu.
                     TAG_SEQ=None found"
BASE COUNT       31 a      77 c      48 g      76 t       1 others
ORIGIN
```

```
  Query Match             0.3%;  Score 20;  DB 138;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     208 TTAATGAAAAAAAAAAAAAA 189


RESULT 135
AI764202/c
LOCUS       AI764202       238 bp    mRNA            EST       25-JUN-1999
DEFINITION  UI-R-Y0-acw-b-02-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-acw-b-02-0-UI 3', mRNA sequence.
ACCESSION   AI764202
VERSION     AI764202.1  GI:5210137
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
```

```
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized Eye library cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .238
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-Y0-acw-b-02-0-UI"
                      /clone_lib="UI-R-Y0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
                      library is a subtracted library derived from an
                      individually-tagged normalized whole-eye (minus the lens)
                      library. The driver for the subtraction consisted of a
                      pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
                      UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
                      3-5 nucleotides present between the Not I site and the
                      oligo-dT track which allows identification of the library
                      of origin of a clone within the mixture. The subtracted
                      library (UI-R-Y0) was constructed as follows: PCR
                      amplified cDNA inserts from previous library clones from
                      which 3' ESTs had  been derived were used as a driver in a
                      hybridization with the normalized  whole-eye library in
                      the form of single-stranded circles. The remaining
                      single-stranded circles (subtracted library) was purified
                      by hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-Y0
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)
                      TAG_LIB=UI-R-Y0
                      TAG_TISSUE=Eye
```



```
                      TAG_SEQ=CATTG"
BASE COUNT       60 a      57 c      43 g      78 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 24;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      23 TTAATGAAAAAAAAAAAAAA 4


RESULT 136
AW070146
LOCUS       AW070146       238 bp    mRNA            EST       13-OCT-1999
DEFINITION  mgiel005D11f Subtracted Rice blast Infection Stage cDNA library
            Oryza sativa/Pyricularia oryzae mixed EST library cDNA clone
            mgiel005D11f, mRNA sequence.
ACCESSION   AW070146
VERSION     AW070146.1  GI:6025144
KEYWORDS    EST.
SOURCE      Oryza sativa/Pyricularia oryzae mixed EST library.
  ORGANISM  Oryza sativa/Pyricularia oryzae mixed EST library
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Asiegbu,P., Choi;W., Rauyaree,P. and Dean,R.A.
  TITLE     Expressed sequence characterization of subtracted genes during
            infection stage of rice blast
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Dean,R.A.
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson Universiy, Clemson, SC 29634, USA
            Tel: 864 656 5737
            Fax: 864 656 4293
            Email: rdean@clemson.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG)
            High quality sequence start: 30
            High quality sequence stop: 102.
FEATURES              Location/Qualifiers
     source           1. .238
                      /organism="Oryza sativa/Pyricularia oryzae mixed EST
                      library"
                      /strain="70-15"
                      /db_xref="taxon:105664"
                      /clone="mgiel005D11f"
                      /clone_lib="Subtracted Rice blast Infection Stage cDNA
                      library"
                      /dev_stage="Infection stage, 48hrs, Subtraction with
```



```
                      healthy plant"
                      /note="Vector: pBlueScriptII SK(+) Vector; Rice blast
                      infection stage mRNA population at 48 hrs after
                      inoculation was subtracted with mRNA population of healthy
                      plant."
BASE COUNT       83 a      43 c      38 g      74 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 39;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db     196 AAAAAAAAAAAAAATGCTTG 215


RESULT 137
BE804542
LOCUS       BE804542       238 bp    mRNA            EST       20-SEP-2000
DEFINITION  sr83a04.y1 Gm-c1047 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1047-7 5', mRNA sequence.
ACCESSION   BE804542
VERSION     BE804542.1  GI:10235654
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
```



```
            info@genomesystems.com web site: www.genomesystems.com.
FEATURES              Location/Qualifiers
     source           1. .238
                      /organism="Glycine max"
                      /db_xref="taxon:3847"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1047-7"
                      /clone_lib="Gm-c1047"
                      /tissue_type="immature leaves (unfurled trifoliate) of
                      greenhouse grown plants"
                      /dev_stage="2 week old"
                      /lab_host="DH10B"
                      /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                      cDNA library was constructed from mRNA isolated from
                      immature leaves (unfurled trifoliate) of greenhouse grown
                      plants that were 2 week old. The library was prepared
                      using the Life Technologies pSuperScript cDNA library
                      construction kit. Complementary DNA was synthesized from
                      mRNA using a poly(dT) sequence with a NotI restrictions
                      site. SalI linkers adapters were ligated to the
                      blunt-ended cDNA fragments followed by NotI digestion. The
                      cDNA fragments were directionally cloned into the
                      NotI-SalI restriction site of the pSPORT1 vector. The
                      ligated cDNA fragments were transformed into E.coli
                      ElectroMax DH10B host cells. This library was constructed
                      in the laboratory of Dr. Lila Vodkin by Anu Khanna at the
                      University of Illinois at Urbana-Champaign (email
                      l-vodkin@uiuc.edu)"
BASE COUNT      111 a      47 c      37 g      43 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 136;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     192 TTAATGAAAAAAAAAAAAAA 211


RESULT 138
AI263668/c
LOCUS       AI263668       240 bp    mRNA            EST       13-NOV-1998
DEFINITION  qx67f09.x1 NCI_CGAP_Ov36 Homo sapiens cDNA clone IMAGE:2006441 3',
            mRNA sequence.
ACCESSION   AI263668
VERSION     AI263668.1  GI:3871871
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: +40UP from Gibco
            High quality sequence stop: 230.
FEATURES              Location/Qualifiers
     source           1. .240
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2006441"
                      /clone_lib="NCI_CGAP_Ov36"
                      /sex="female"
                      /tissue_type="borderline ovarian carcinoma"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: ovary; Vector: pAMP1; mRNA made from
                      borderline ovarian carcinoma, cDNA made by oligo-dT
                      priming.  Directionally cloned.  Size-selected on agarose
                      gel, average insert size 500 bp.  Primary library,
                      non-amplified. "
BASE COUNT       82 a      24 c      31 g     103 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 18;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     32 TTAATGAAAAAAAAAAAAAA 13


RESULT 139
AU073502
LOCUS       AU073502      240 bp    mRNA            EST       24-JUN-1999
DEFINITION  AU073502 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSH383, mRNA sequence.
ACCESSION   AU073502
```



```
VERSION     AU073502.1  GI:5179923
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum (1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES              Location/Qualifiers
     source           1. .240
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="SSH383"
                      /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                      /dev_stage="slug"
BASE COUNT      131 a      14 c      24 g      71 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 29;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3172 aataacaataacaataatgg 3191
          ||||||||||||||||||||
Db     48 AATAACAATAACAATAATGG 67


RESULT 140
AU074261
LOCUS       AU074261      240 bp    mRNA            EST       24-JUN-1999
DEFINITION  AU074261 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSK170, mRNA sequence.
ACCESSION   AU074261
VERSION     AU074261.1  GI:5180682
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum (1999)
  JOURNAL   Unpublished (1999)
```



```
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES              Location/Qualifiers
     source           1. .240
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="SSK170"
                      /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                      /dev_stage="slug"
BASE COUNT      131 a      16 c       7 g      86 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 29;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3170 tcaataacaataacaataat 3189
          ||||||||||||||||||||
Db      2 TCAATAACAATAACAATAAT 21


RESULT 141
Z21539/c
LOCUS       Z21539        242 bp    mRNA            EST       01-MAY-1998
DEFINITION  HSAFIB007 Stratagene cDNA library Human fibroblast, cat#937212 Homo
            sapiens cDNA clone Fib007, mRNA sequence.
ACCESSION   Z21539
VERSION     Z21539.1  GI:28456
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Contact: Genzentrum Muenchen
            Laboratorium fuer molekulare Biologie
            Am Klopferspitz 18a,8033 Martinsried,Germany
```



```
            Email: obermaier@vms.biochem.mpg.de
            single read;
             Genexpress library reference is A;.
FEATURES              Location/Qualifiers
     source           1. .242
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Fib007"
                      /clone_lib="Stratagene cDNA library Human fibroblast,
                      cat#937212"
                      /note="Vector: pBluescript SK(+); Genexpress library
                      reference is A. "
BASE COUNT       84 a      37 c      25 g      94 t      2 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 147;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     23 TTAATGAAAAAAAAAAAAAA 4


RESULT 142
AI623688/c
LOCUS       AI623688      244 bp    mRNA            EST       14-DEC-1999
DEFINITION  ts22h10.xl NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2229379 3',
            mRNA sequence.
ACCESSION   AI623688
VERSION     AI623688.1  GI:4648619
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 2086   Std Error: 0.00
```

```
            Seq primer: -40UP from Gibco
            High quality sequence stop: 190
            POLYA=No.
  FEATURES              Location/Qualifiers
       source           1. .244
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:2229379"
                        /clone_lib="NCI_CGAP_Pan1"
                        /tissue_type="adenocarcinoma"
                        /lab_host="DH10B"
                        /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                        Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                        Average insert size 1.72 kb. Life Technologies catalog #:
                        11548-013"
  BASE COUNT       69 a     38 c     31 g    106 t
  ORIGIN


    Query Match             0.3%;  Score 20;  DB 22;  Length 244;
    Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
    Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy    383 ttaatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||||
  Db     21 TTAATGAAAAAAAAAAAAAA 2


  RESULT 143
  AV021342/c
  LOCUS       AV021342      244 bp    mRNA            EST       28-AUG-1999
  DEFINITION  AV021342 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
              clone 1190022I10, mRNA sequence.
  ACCESSION   AV021342
  VERSION     AV021342.1  GI:4798334
  KEYWORDS    EST.
  SOURCE      house mouse.
    ORGANISM  Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE   1  (bases 1 to 244)
    AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
              Akahira,S., Akiyama,J., Fukuda,S., Fukuishi,Y., Funayama,T., Hara
              ,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
              Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,
              Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
              ,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,
              Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
              Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
    TITLE     RIKEN House ESTs
    JOURNAL   Unpublished (1999)
```



```
  COMMENT     Contact: Chie Owa
              Genome Science Laboratory
              RIKEN
              3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
              Tel: 81-298-36-9145
              Fax: 81-298-36-9098
              Email: genome-res@rtc.riken.go.jp
              Thermostabilization and thermoactivation of thermolabile enzymes by
              trehalose and its application for the synthesis of full length cDNA
              (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
              Transcriptional sequencing: A method for DNA sequencing using RNA
              polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
              Please visit our web site (http://genome.rtc.riken.go.jp) for
              further details.
  FEATURES              Location/Qualifiers
       source           1. .244
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="1190022I10"
                        /clone_lib="Mus musculus 18-day embryo C57BL/6J"
                        /sex="mixed"
                        /dev_stage="18-day embryo"
  BASE COUNT       43 a     31 c     39 g    131 t
  ORIGIN


    Query Match             0.3%;  Score 20;  DB 29;  Length 244;
    Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
    Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy    383 ttaatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||||
  Db    230 TTAATGAAAAAAAAAAAAAA 211


  RESULT 144
  AI974381
  LOCUS       AI974381      247 bp    mRNA            EST       26-AUG-1999
  DEFINITION  T110323e KV0 Medicago truncatula cDNA clone pKV0-1N4, mRNA
              sequence.
  ACCESSION   AI974381
  VERSION     AI974381.1  GI:5776762
  KEYWORDS    EST.
  SOURCE      barrel medic.
    ORGANISM  Medicago truncatula
              Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
              Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
              Fabales; Fabaceae; Papilionoideae; Medicago.
  REFERENCE   1  (bases 1 to 247)
    AUTHORS   VandenBosch,K., Hur,J., Moore,J., Beremand,P., Peng,H. and Ellis,L.
```



```
    TITLE     ESTs from uninoculated roots of Medicago truncatula (1999b)
    JOURNAL   Unpublished (1999)
  COMMENT     Contact: VandenBosch K
              Department of Biology
              Texas A&M University
              College Station, TX 77843-3258, USA
              Tel: 409 845 7707
              Fax: 409 845 2891
              Email: kate@mail.bio.tamu.edu
              Other name: 08-D-2; date: 7/9/99; Submitted to the Database of
              Expressed Sequence Tags (dbEST) on 08/25/99; More information is
              available at 'http://chrysie.tamu.edu/medicago'.
              Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
  FEATURES              Location/Qualifiers
       source           1. .247
                        /organism="Medicago truncatula"
                        /cultivar="genotype A17"
                        /db_xref="taxon:3880"
                        /clone="pKV0-1N4"
                        /clone_lib="KV0"
                        /tissue_type="Seedling roots"
                        /dev_stage="Immediately prior to inoculation with
                        Sinorhizobium meliloti (0 hour)"
                        /lab_host="E.coli strain XLOLR"
                        /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                        XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                        was directionally ligated into the Unizap XR vector from
                        Stratagene and packaged using Gigapack III Gold packaging
                        extracts. Plasmids containing cDNA inserts were excised
                        from the recombinant lambda-Zap phage using Ex-assist
                        helper phage and propagated in XLOLR cells."
  BASE COUNT      109 a     52 c     33 g     53 t
  ORIGIN


    Query Match             0.3%;  Score 20;  DB 27;  Length 247;
    Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
    Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy   2473 caagctaacagcccaaaaaga 2492
            ||||||||||||||||||||
  Db    163 CAAGCTAACAGCCAAAAAGA 182


  RESULT 145
  AA579530
  LOCUS       AA579530      252 bp    mRNA            EST       12-SEP-1997
  DEFINITION  nf37h06.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:915995, mRNA
              sequence.
  ACCESSION   AA579530
  VERSION     AA579530.1  GI:2357714
```



```
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 252)
    AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
    JOURNAL   Unpublished (1997)
  COMMENT     Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert_Strausberg@nih.gov
              Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
              , Michael Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: David B. Krizman, Ph.D.
              cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert Length: 252  Std Error: 0.00
              Seq primer: -40m13 fwd. ET from Amersham.
  FEATURES              Location/Qualifiers
       source           1. .252
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:915995"
                        /clone_lib="NCI_CGAP_Pr2"
                        /sex="Male"
                        /dev_stage="45 years old"
                        /lab_host="DH10B"
                        /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                        strand cDNA was primed with oligo(dT)17 on 50 ng of
                        DNAse-treated, total cellular RNA obtained from 5,000-10
                        ,000 microdissected preneoplastic cells
                        histologically-determined to be prostatic intraepithelial
                        neoplasia 2 (PIN2) cells.  Double-stranded cDNA was
                        ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                        cDNA with an adaptor-specific primer, and the resulting
                        PCR product subcloned into pAMP10 by the UDG-cloning
                        method (Life Technologies).  Average insert size is 600
                        bp.  NOTE: Not directionally cloned.  This library was
                        constructed by  David Krizman."
  BASE COUNT      116 a     27 c     37 g     72 t
  ORIGIN


    Query Match             0.3%;  Score 20;  DB 9;  Length 252;
    Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
    Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     215 TTAATGAAAAAAAAAAAAAA 234


RESULT 146
AW109600/c
LOCUS       AW109600      253 bp    mRNA            EST       20-OCT-1999
DEFINITION  gate0004N20f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone gate0004N20f, mRNA sequence.
ACCESSION   AW109600
VERSION     AW109600.1  GI:6080400
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Routh,G., Frisch,D., Woo,S.S., Wood,T.C., Wing,R.A. and Wilkins
            ,T.A.
  TITLE     Generation and analysis of 1250 EST's derived from rapidly growing
            diploid cotton fiber cDNAs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 215.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="gate0004N20f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       91 a      24 c      66 g      72 t
ORIGIN

Query Match              0.3%;  Score 20;  DB 40;  Length 253;
```



```
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2588 ataaacttgaaaatgtcaac 2607
           ||||||||||||||||||||
Db     188 ATAAACTTGAAAATGTCAAC 169


RESULT 147
AL368864
LOCUS       AL368864      254 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBA27C04R1 MtBA Medicago truncatula cDNA clone MtBA27C04 T7, mRNA
            sequence.
ACCESSION   AL368864
VERSION     AL368864.1  GI:9668617
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from nitrogen-starved roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .254
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBA27C04"
                     /clone_lib="MtBA"
                     /tissue_type="root tips"
                     /dev_stage="harvested after 3 days of N-starvation"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber for 14
                     days on nitrogen-rich medium followed by 3 days on N-free
                     medium. RNA was extracted from root tips (1-3 cm). cDNA
                     was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zapXR vector from
```



```
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT       77 a      32 c      58 g      87 t
ORIGIN

  Query Match              0.3%;  Score 20;  DB 28;  Length 254;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     215 TTAATGAAAAAAAAAAAAAA 234


RESULT 148
AU058025
LOCUS       AU058025      260 bp    mRNA            EST       29-APR-1999
DEFINITION  AU058025 Rice callus Oryza sativa cDNA clone C60924_1A, mRNA
            sequence.
ACCESSION   AU058025
VERSION     AU058025.1  GI:4714058
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from callus (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .260
                     /organism="Oryza sativa"
                     /strain="cultivar Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="C60924_1A"
```



```
                     /clone_lib="Rice callus"
                     /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                     NotI; cDNA prepared from rice callus mRNAs by using
                     oligo(dT) as a primer and ligating to the SalI-NotI site
                     of pBluescript II SK+ phagemid."
BASE COUNT       86 a      53 c      51 g      70 t
ORIGIN

  Query Match              0.3%;  Score 20;  DB 29;  Length 260;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     233 TTAATGAAAAAAAAAAAAAA 252


RESULT 149
AA788980/c
LOCUS       AA788980      261 bp    mRNA            EST       12-JAN-1999
DEFINITION  ah29b12.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            1240223 3' similar to contains element PTR5 repetitive element ;,
            mRNA sequence.
ACCESSION   AA788980
VERSION     AA788980.1  GI:2849100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1142  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 188.
FEATURES             Location/Qualifiers
     source          1. .261
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="1240223"
/clone_lib="Soares_parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTT TTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 89 a 27 c 38 g 107 t
ORIGIN

Query Match 0.3%; Score 20; DB 12; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAAA 4

RESULT 150
AI481145/c
LOCUS AI481145 262 bp mRNA EST 09-MAR-1999
DEFINITION vg86e04.x1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA clone IMAGE:872862 3', mRNA sequence.
ACCESSION AI481145
VERSION AI481145.1 GI:4374371
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 262)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:512342
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
High quality sequence stop: 252.
FEATURES Location/Qualifiers
source 1. .262
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:872862"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTT 3'); double-stranded cDNA was ligated to Eco RI adaptors (GTTGGATTCGGTACC), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."
BASE COUNT 69 a 49 c 61 g 83 t
ORIGIN

Query Match 0.3%; Score 20; DB 21; Length 262;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 45 TTAATGAAAAAAAAAAAAAA 26

RESULT 151
AW490767/c
LOCUS AW490767 262 bp mRNA EST 24-FEB-2000
DEFINITION UI-M-BH3-asx-h-05-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone

UI-M-BH3-asx-h-05-0-UI 3', mRNA sequence.
ACCESSION AW490767
VERSION AW490767.1 GI:7061040
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 262)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized pineal glands library cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .262
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-asx-h-05-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the

representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=pineal-glands
TAG_SEQ=CAGAC"
BASE COUNT 77 a 45 c 47 g 93 t
ORIGIN

Query Match 0.3%; Score 20; DB 91; Length 262;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAAA 4

RESULT 152
BE052081/c
LOCUS BE052081 262 bp mRNA EST 08-JUN-2000
DEFINITION GA__Ea0004N20f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA__Ea0004N20f, mRNA sequence.
ACCESSION BE052081
VERSION BE052081.1 GI:8379137
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 262)
AUTHORS Leslie,A., Frisch,D., Yu,Y., Wood,T.C., Wing,R.A. and Wilkins,T.A.
TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson Univerasity Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 16
            High quality sequence stop: 246.
FEATURES             Location/Qualifiers
     source          1. .262
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0004N20f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       94 a      26 c      70 g      72 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 134;  Length 262;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2588 ataaacttgaaaatgtcaac 2607
           ||||||||||||||||||||
Db     196 ATAAACTTGAAAATGTCAAC 177


RESULT 153
AI250156/c
LOCUS       AI250156      265 bp    mRNA            EST       21-DEC-1998
DEFINITION  qx49a05.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2004656 3',
            mRNA sequence.
ACCESSION   AI250156
VERSION     AI250156.1  GI:3846685
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```



```
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Insert Length: 1448    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 263.
FEATURES             Location/Qualifiers
     source          1. .265
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2004656"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT      113 a      40 c      28 g      83 t       1 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 18;  Length 265;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      23 TTAATGAAAAAAAAAAAAAA 4


RESULT 154
BE953755/c
LOCUS       BE953755      273 bp    mRNA            EST       04-OCT-2000
DEFINITION  UI-M-CC1-azx-f-07-0-UI.s1 NIH_BMAP_Ret1_N Mus musculus cDNA clone
            UI-M-CC1-azx-f-07-0-UI 3', mRNA sequence.
ACCESSION   BE953755
VERSION     BE953755.1  GI:10595581
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
```



```
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            embryonic retina tissue cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: Researchers may obtain BMAP cDNA clones from
            RESEARCH GENETICS. It should be noted that Bento Soares is
            generating a small number of additional specialized non-redundant
            arrays of BMAP cDNAs whose availability will be considered under
            appropriate and limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .273
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-CC1-azx-f-07-0-UI"
                     /clone_lib="NIH_BMAP_Ret1_N"
                     /dev_stage="13.5 days pc"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret1_N library is a normalized library derived
                     NIH_BMAP_Ret1. NIH_BMAP_Ret1 was made from  mouse
                     embryonic retina tissue. For a detailed description of the
                     library from which this clone was  derived, please visit
                     our web site at brainest.eng.uiowa.edu.
                     TAG_LIB=NIH_BMAP_Ret1_N
                     TAG_TISSUE=embryonic-retina
                     TAG_SEQ=CTGCTAGTAGCA"
BASE COUNT       65 a      40 c      54 g     114 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 138;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      22 TTAATGAAAAAAAAAAAAAA 3
```



```
RESULT 155
AI561501/c
LOCUS       AI561501      277 bp    mRNA            EST       25-MAR-1999
DEFINITION  vw88c12.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            IMAGE:1262038 3', mRNA sequence.
ACCESSION   AI561501
VERSION     AI561501.1  GI:4512846
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 277)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:664590
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 225.
FEATURES             Location/Qualifiers
     source          1. .277
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1262038"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
```

```
BASE COUNT       70 a     40 c     54 g    113 t
ORIGIN

  Query Match             0.3%;  Score 20;  DB 22;  Length 277;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     35 TTAATGAAAAAAAAAAAAAA 16


RESULT 156
AQ976175/c
LOCUS       AQ976175      283 bp    DNA             GSS       28-JAN-2000
DEFINITION  RPCI-23-333N8.TJ RPCI-23 Mus musculus genomic clone RPCI-23-333N8,
            DNA sequence.
ACCESSION   AQ976175
VERSION     AQ976175.1  GI:6806632
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 283)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 333  row: N  column: 8
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .283
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
```



```
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-333N8"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT       71 a     75 c     34 g    102 t      1 others
ORIGIN

  Query Match             0.3%;  Score 20;  DB 173;  Length 283;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db    160 TTAATGAAAAAAAAAAAAAA 141


RESULT 157
BB366530/c
LOCUS       BB366530      284 bp    mRNA            EST       12-JUL-2000
DEFINITION  BB366530 RIKEN full-length enriched, 16 days embryo head Mus
            musculus cDNA clone C130034B02 3', mRNA sequence.
ACCESSION   BB366530
VERSION     BB366530.1  GI:9078358
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 284)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
```



```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .284
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C130034B02"
                     /clone_lib="RIKEN full-length enriched, 16 days embryo
                     head"
                     /sex="mixed"
                     /tissue_type="head"
                     /dev_stage="16 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
```



```
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I"
BASE COUNT      112 a     51 c     40 g     81 t
ORIGIN

  Query Match             0.3%;  Score 20;  DB 128;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5822 ttttttaatttaatgggatt 5841
          ||||||||||||||||||||
Db     70 TTTTTTAATTTAATGGGATT 51


RESULT 158
AU036991
LOCUS       AU036991      285 bp    mRNA            EST       29-MAR-1999
DEFINITION  AU036991 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSA714, mRNA sequence.
ACCESSION   AU036991
VERSION     AU036991.1  GI:3983744
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 285)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .285
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSA714"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      161 a     47 c     14 g     63 t
```

ORIGIN

Query Match 0.3%; Score 20; DB 29; Length 285;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3170 tcaataacaataacaataat 3189
||||||||||||||||||||
Db 52 TCAATAACAATAACAATAAT 71

RESULT 159
BB265787
LOCUS BB265787 285 bp mRNA EST 07-JUL-2000
DEFINITION BB265787 RIKEN full-length enriched, 10 days neonate cortex Mus musculus cDNA clone A830022C19 3', mRNA sequence.
ACCESSION BB265787
VERSION BB265787.1 GI:8962248
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 285)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1..285
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A830022C19"
/clone_lib="RIKEN full-length enriched, 10 days neonate cortex"
/tissue_type="cortex"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 94 a 40 c 48 g 103 t
ORIGIN

Query Match 0.3%; Score 20; DB 126; Length 285;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 aaaaaaatgcttgatattgt 415
||||||||||||||||||||
Db 89 AAAAAAATGCTTGATATTGT 108

RESULT 160
BB264068
LOCUS BB264068 287 bp mRNA EST 07-JUL-2000
DEFINITION BB264068 RIKEN full-length enriched, 10 days neonate cortex Mus musculus cDNA clone A830014A04 3', mRNA sequence.
ACCESSION BB264068
VERSION BB264068.1 GI:8960524
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 287)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1..287
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A830014A04"
/clone_lib="RIKEN full-length enriched, 10 days neonate cortex"
/tissue_type="cortex"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 95 a 41 c 49 g 102 t
ORIGIN

Query Match 0.3%; Score 20; DB 126; Length 287;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 aaaaaaatgcttgatattgt 415
||||||||||||||||||||
Db 91 AAAAAAATGCTTGATATTGT 110

RESULT 161
BB246243
LOCUS BB246243 291 bp mRNA EST 06-JUL-2000
DEFINITION BB246243 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730013M08 3', mRNA sequence.
ACCESSION BB246243
VERSION BB246243.1 GI:8938989
KEYWORDS EST.

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES                Location/Qualifiers
     source             1. .291
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /clone="A730013M08"
                        /clone_lib="RIKEN full-length enriched, 7 days neonate
                        cerebellum"
```



```
                        /tissue_type="cerebellum"
                        /dev_stage="7 days neonate"
                        /lab_host="DH10B"
                        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap-trapper. cDNA went through one round of normalization
                        to Rot = 20.0 and subtraction to Rot = 459.0. Second
                        strand cDNA was prepared with the primer adapter of
                        sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                        3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                        modified pBluescript KS(+) after bulk excision from Lambda
                        FLC I."
BASE COUNT       57 a     76 c     61 g     97 t
ORIGIN
```

```
  Query Match             0.3%;  Score 20;  DB 103;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      75 tttgttttccattttgtttc 94
           ||||||||||||||||||||
Db     236 TTTGTTTTCCATTTTGTTTC 255
```

```
RESULT 162
AI835732/c
LOCUS       AI835732      292 bp    mRNA            EST       14-JUL-1999
DEFINITION  UI-M-AI0-aan-e-02-0-UI.s1 NIH_BMAP_MBS Mus musculus cDNA clone
            UI-M-AI0-aan-e-02-0-UI 3', mRNA sequence.
ACCESSION   AI835732
VERSION     AI835732.1  GI:5469945
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
```



```
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized brain stems library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES                Location/Qualifiers
     source             1. .292
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="UI-M-AI0-aan-e-02-0-UI"
                        /clone_lib="NIH_BMAP_MBS"
                        /dev_stage="27-32 days"
                        /lab_host="DH10B (Life Technologies)"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; The
                        NIH_BMAP_MBS library is a non-normalized library
                        constructed from  mouse brain stems. The tag is a string
                        of 5 nucleotides present  between the Not I site and the
                        oligo-dT track.  The library was  constructed as described
                        by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                        , 1996. Tissue provided by Ms. Annie Novakovich,
                        Zivic-Miller Laboratories.
                        TAG_LIB=NIH_BMAP_MBS
                        TAG_TISSUE=brain-stems
                        TAG_SEQ=TCANG"
BASE COUNT       72 a     61 c     74 g     85 t
ORIGIN
```

```
  Query Match             0.3%;  Score 20;  DB 25;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      23 TTAATGAAAAAAAAAAAAAA 4
```



```
RESULT 163
AI602026/c
LOCUS       AI602026      294 bp    mRNA            EST       21-APR-1999
DEFINITION  UI-R-Y0-uv-h-10-0-UI.s2 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-uv-h-10-0-UI 3', mRNA sequence.
ACCESSION   AI602026
VERSION     AI602026.1  GI:4611187
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized Eye library cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.resgen.com) The following repetitive elements were
            found in this cDNA sequence: 1-27, >AT_rich#Low_complexity
            Seq primer: M13 Forward.
FEATURES                Location/Qualifiers
     source             1. .294
                        /organism="Rattus norvegicus"
                        /strain="Sprague-Dawley"
                        /db_xref="taxon:10116"
                        /clone="UI-R-Y0-uv-h-10-0-UI"
                        /clone_lib="UI-R-Y0"
                        /dev_stage="adult"
                        /lab_host="DH10B (Life Technologies)"
                        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
                        library is a subtracted library derived from an
                        individually-tagged normalized whole-eye (minus the lens)
```

library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
BASE COUNT 63 a 83 c 63 g 84 t 1 others
ORIGIN

Query Match 0.3%; Score 20; DB 22; Length 294;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAA 4

RESULT 164
AI661433/c
LOCUS AI661433 294 bp mRNA EST 10-MAY-1999
DEFINITION va26e12.x1 GuayWoodford Beier mouse kidney day 7 Mus musculus cDNA clone IMAGE:732526 3', mRNA sequence.
ACCESSION AI661433
VERSION AI661433.1 GI:4765016
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 294)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)

COMMENT Other_ESTs: va26e12.y1
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:449574
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
Seq primer: Primer name ambiguous
High quality sequence stop: 151.
FEATURES Location/Qualifiers
source 1. .294
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:732526"
/clone_lib="GuayWoodford Beier mouse kidney day 7"
/tissue_type="kidney"
/dev_stage="juvenile (7 days old)"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: kidney; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3' Library provided by Lisa Guay-Woodford."
BASE COUNT 95 a 66 c 43 g 90 t
ORIGIN

Query Match 0.3%; Score 20; DB 23; Length 294;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 21 TTAATGAAAAAAAAAAAAA 2

RESULT 165
AW492508/c
LOCUS AW492508 294 bp mRNA EST 24-FEB-2000
DEFINITION UI-M-BH3-atg-c-12-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone UI-M-BH3-atg-c-12-0-UI 3', mRNA sequence.
ACCESSION AW492508
VERSION AW492508.1 GI:7062789
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 294)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized basal ganglia library cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements The following repetitive elements were found in this cDNA sequence: 1-44, >AT_rich#Low_complexity 205-226, >AT_rich#Low_complexity
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .294
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-atg-c-12-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries

were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=basal-ganglia
TAG_SEQ=TGTAC"
BASE COUNT 101 a 33 c 58 g 102 t
ORIGIN

Query Match 0.3%; Score 20; DB 91; Length 294;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAA 4

RESULT 166
BE116692/c
LOCUS BE116692 294 bp mRNA EST 13-JUN-2000
DEFINITION UI-R-BS1-aye-b-11-0-UI.s1 UI-R-BS1 Rattus norvegicus cDNA clone UI-R-BS1-aye-b-11-0-UI 3', mRNA sequence.
ACCESSION BE116692
VERSION BE116692.1 GI:8508797
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 294)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477

```
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized embryo at 13 dpc library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 1-54,
            >AT_rich=Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .294
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-BS1-aye-b-11-0-UI"
                      /clone_lib="UI-R-BS1"
                      /dev_stage="embryonic 13 dpc"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS1
                      library is derived from 13 dpc whole embryo tissue.  For a
                      detailed description of the library from which this clone
                      was derived, please visit our web site at
                      ratest.eng.uiowa.edu.
                      TAG_LIB=UI-R-BS1
                      TAG_TISSUE=embryo at 13 dpc
                      TAG_SEQ=AATCC"
BASE COUNT       87 a      60 c      38 g     109 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 104;  Length 294;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      23 TTAATGAAAAAAAAAAAAAA 4


RESULT 167
```



```
W10380
LOCUS       W10380        295 bp    mRNA            EST       01-OCT-1997
DEFINITION  ma38c07.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:312972 5', mRNA sequence.
ACCESSION   W10380
VERSION     W10380.1  GI:1284697
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 295)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI: 203588
            Seq primer: mob.REGA+ET
            High quality sequence stop: 231.
FEATURES              Location/Qualifiers
     source           1. .295
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:312972"
                      /clone_lib="Soares mouse p3NMF19.5"
                      /dev_stage="19.5 dpc total fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Vector: pT7T3D (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. RNA was kindly provided by
                      Dr. Minoru Ko (Wayne State University)."
BASE COUNT       86 a      85 c      79 g      45 t
```



```
ORIGIN


  Query Match             0.3%;  Score 20;  DB 146;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      29 TTAATGAAAAAAAAAAAAAA 48


RESULT 168
AI084635/c
LOCUS       AI084635      298 bp    mRNA            EST       01-OCT-1998
DEFINITION  or78d04.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
            clone IMAGE:1681447 3', mRNA sequence.
ACCESSION   AI084635
VERSION     AI084635.1  GI:3423058
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 298)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1662  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 231.
FEATURES              Location/Qualifiers
     source           1. .298
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1681447"
                      /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                      /tissue_type="senescent fibroblast"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Vector: pT7T3D (Pharmacia) with a modified
                      polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI
                      ; 1st strand cDNA was primed with a Not I - oligo(dT)
                      primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
```



```
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo."
BASE COUNT       89 a      57 c      76 g      76 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 16;  Length 298;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      22 TTAATGAAAAAAAAAAAAAA 3


RESULT 169
AU077610
LOCUS       AU077610      298 bp    mRNA            EST       30-JUL-1999
DEFINITION  AU077610 Rice panicle (longer than 10cm) Oryza sativa cDNA clone
            E20630_8A, mRNA sequence.
ACCESSION   AU077610
VERSION     AU077610.1  GI:5667368
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 298)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from panicle (longer than 10cm)
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES              Location/Qualifiers
     source           1. .298
                      /organism="Oryza sativa"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:4530"
                      /clone="E20630_8A"
                      /clone_lib="Rice panicle (longer than 10cm)"
```

```
                     /tissue_type="panicle"
                     /dev_stage="panicle (longer than 10cm)"
BASE COUNT       98 a      65 c      65 g      70 t
ORIGIN

  Query Match             0.3%;  Score 20;  DB 29;  Length 298;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     255 TTAATGAAAAAAAAAAAAAA 274


RESULT 170
AU057893
LOCUS       AU057893      299 bp    mRNA            EST       29-APR-1999
DEFINITION  AU057893 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA
            clone S21882_2Z, mRNA sequence.
ACCESSION   AU057893
VERSION     AU057893.1  GI:4716777
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from mature leaf
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="S21882_2Z"
                     /clone_lib="Oryza sativa mature leaf Nipponbare"
                     /tissue_type="mature leaf"
BASE COUNT       85 a      38 c      45 g     128 t       3 others
ORIGIN
```



```
  Query Match             0.3%;  Score 20;  DB 29;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     267 TTAATGAAAAAAAAAAAAAA 286


RESULT 171
AW016180/c
LOCUS       AW016180      299 bp    mRNA            EST       10-SEP-1999
DEFINITION  UI-H-BI0p-abx-d-01-0-UI.s1 NCI_CGAP_Sub2 Homo sapiens cDNA clone
            IMAGE:2712216 3', mRNA sequence.
ACCESSION   AW016180
VERSION     AW016180.1  GI:5864937
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html The following repetitive
            elements were found in this cDNA sequence: 1-37,
            >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2712216"
                     /clone_lib="NCI_CGAP_Sub2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
```



```
                     NCI_CGAP_Sub2 library is a subtracted library derived from
                     BI. BI constitutes a mixture of 21 normalized or
                     subtracted NCI_CGAP libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22
                     , NCI_CGAP_Pr28, NCI_CGAP_Co10, NCI_CGAP_Co16,
                     NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3,
                     NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8,
                     NCI_CGAP_CLL1, NCI_CGAP_Lei2, NCI_CGAP_Brn23, NCI_CGAP_Lu5
                     , NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6
                     , NCI_CGAP_Brn25.  These 21 libraries were pooled and a
                     single-stranded DNA preparation of the resulting mixture
                     was used as a tracer in a subtractive hybridization with a
                     driver whose composition is detailed below: NCI_CGAP_Kid3
                     pool 1 LLAM 3334-3337, 3682-3683, 3798-3803   (IMAGE
                     CloneIDs 1322376-1323911, 1456008-1456775, 1500552-1502855
                     ) NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725,
                     3776-3778   (IMAGE CloneIDs 1323912-1325831,
                     1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1 LLAM
                     3575-3582, 3851-3854   (IMAGE CloneIDs 1414920-1417991,
                     1520904-1522439)  NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
                     3716-3720, 3733-3735   (IMAGE CloneIDs 1257096-1258631,
                     1469064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1
                     LLAM 2457-2459, 2758-2759, 3062-3068   (IMAGE CloneIDs
                     985608-986759, 1101192-1101959, 1217928-1220615)
                     NCI_CGAP_Co10 pool 1 LLAM 2644-2651, 2871-2872   (IMAGE
                     CloneIDs 1057416-1061255, 1144584-1145351)  The resulting
                     subtracted library contained 4 million recombinants.
                     Subtraction was performed as previously described [Bonaldo
                     ,  Lennon & Soares (1996): Normalization and Subtraction:
                     Two Approaches  To Facilitate Gene Discovery. Genome
                     Research 6, 791-806.
                     TAG_LIB=NCI_CGAP_GC4
                     TAG_TISSUE=germ cell
                     TAG_SEQ=AAATC"
BASE COUNT       89 a      57 c      47 g     106 t
ORIGIN

  Query Match             0.3%;  Score 20;  DB 38;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      23 TTAATGAAAAAAAAAAAAAA 4


RESULT 172
BE108199
LOCUS       BE108199      299 bp    mRNA            EST       13-JUN-2000
DEFINITION  UI-R-BS1-ayw-b-09-0-UI.s1 UI-R-BS1 Rattus norvegicus cDNA clone
            UI-R-BS1-ayw-b-09-0-UI 3', mRNA sequence.
```



```
ACCESSION   BE108199
VERSION     BE108199.1  GI:8500269
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail.  The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized embryo at 13 dpc library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 226-281,
            >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BS1-ayw-b-09-0-UI"
                     /clone_lib="UI-R-BS1"
                     /dev_stage="embryonic 13 dpc"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS1
                     library is derived from 13 dpc whole embryo tissue.  For a
                     detailed description of the library from which this clone
                     was derived, please visit our web site at
                     ratest.eng.uiowa.edu.
                     TAG_LIB=UI-R-BS1
                     TAG_TISSUE=embryo at 13 dpc
                     TAG_SEQ=AATCC"
```

```
BASE COUNT      108 a     40 c     57 g     93 t      1 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 104;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db     257 TTAATGAAAAAAAAAAAAA 276


RESULT 173
B05848
LOCUS       B05848          301 bp    DNA             GSS       13-JUL-1996
DEFINITION  cSRL-72a5-u cSRL flow sorted Chromosome 11 specific cosmid Homo
            sapiens genomic clone cSRL-72a5, DNA sequence.
ACCESSION   B05848
VERSION     B05848.1  GI:1415126
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 301)
  AUTHORS   Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
            Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S. ., Harris
            ,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and
            Garner,H.R.
  TITLE     Genomic Sequence Sampled Map of Chromosome 11
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Evans GA, Shane Probst
            McDermott Center for Human Growth and Development
            University of Texas Southwestern Medical Center At Dallas
            5323 Harry Hines Blvd, Dallas TX 75235-8591
            Tel: 214-648-1600
            Fax: 214-648-1666
            Email: gevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
            PCR PRimers
            FORWARD: AATATCAGGCATCTTTTGTG
            BACKWARD: TTTGAATGGTTTTGGTAAAG
            Seq primer: T7
            Class: cosmid ends
            High quality sequence stop: 301.
FEATURES             Location/Qualifiers
     source          1. .301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="cSRL-72a5"
                     /clone_lib="cSRL flow sorted Chromosome 11 specific
```



```
                     cosmid"
                     /sex="female"
                     /cell_type="chimeric hamster somatic cell hybrid"
                     /note="Vector: sCos-1; Human Chromosome 11 specific cosmid
                     library prepared from flow sorted human  Chromosome 11
                     derived from Chinese Hampster Ovary (CHO) monochromosomal
                     somatic cell hybrid, J1"
BASE COUNT      117 a     43 c     57 g     69 t     15 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 183;  Length 301;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db     193 AAAAAAAAAAAAAATGCTTG 212


RESULT 174
AI842938/c
LOCUS       AI842938        303 bp    mRNA            EST       14-JUL-1999
DEFINITION  UI-M-AO1-aem-d-04-0-UI.s1 NIH_BMAP_MPG_N Mus musculus cDNA clone
            UI-M-AO1-aem-d-04-0-UI 3', mRNA sequence.
ACCESSION   AI842938
VERSION     AI842938.1  GI:5477151
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 303)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized pineal glands library cDNA Library Preparation: M.B.
```



```
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .303
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-AO1-aem-d-04-0-UI"
                     /clone_lib="NIH_BMAP_MPG_N"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MPG_N library is a normalized library constructed
                     from  mouse pineal gland. The tag is a string of 5
                     nucleotides present  between the Not I site and the
                     oligo-dT track.  The library was  constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996. Tissue provided by Ms. Annie Novakovich,
                     Zivic-Miller Laboratories.
                     TAG_LIB=NIH_BMAP_MPG_N
                     TAG_TISSUE=pineal-glands
                     TAG_SEQ=CAGAC"
BASE COUNT       86 a     52 c     45 g    120 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 25;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      23 TTAATGAAAAAAAAAAAAA 4


RESULT 175
AA669251/c
LOCUS       AA669251        305 bp    mRNA            EST       20-NOV-1997
DEFINITION  ac17e11.s1 Stratagene ovary (#937217) Homo sapiens cDNA clone
            IMAGE:856748 3' similar to contains Alu repetitive element;contains
            element L1 repetitive element ;, mRNA sequence.
ACCESSION   AA669251
VERSION     AA669251.1  GI:2630750
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```



```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 305)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kriman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 297.
FEATURES             Location/Qualifiers
     source          1. .305
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:856748"
                     /clone_lib="Stratagene ovary (#937217)"
                     /sex="female"
                     /dev_stage="49 year old"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
                     Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                     Total ovary tissue, normal, caucasian. Average insert
                     size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT       57 a     89 c     75 g     84 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 10;  Length 305;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5754 ttttaaaaggctttttaaaa 5773
           ||||||||||||||||||||
Db     289 TTTTAAAAGGCTTTTTAAAA 270


RESULT 176
R46662/c
LOCUS       R46662          305 bp    mRNA            EST       22-MAY-1995
DEFINITION  yg53h05.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
```

```
            IMAGE:36684 3', mRNA sequence.
ACCESSION   R46662
VERSION     R46662.1  GI:822601
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 305)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 10, 1995 this sequence version replaced gi:806059.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1531
            High quality sequence stops: 219 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1531  Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 219.
FEATURES             Location/Qualifiers
     source          1. .305
                     /organism="Homo sapiens"
                     /db_xref="GDB:409185"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:36684"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      97 a      44 c      44 g     112 t       8 others
ORIGIN
```



```
  Query Match             0.3%;  Score 20;  DB 144;  Length 305;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      22 TTAATGAAAAAAAAAAAAAA 3


RESULT 177
AQ230212/c
LOCUS       AQ230212      305 bp     DNA             GSS       26-SEP-1998
DEFINITION  HS_2041_B1_F03_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2041 Col=5 Row=L, DNA sequence.
ACCESSION   AQ230212
VERSION     AQ230212.1  GI:3655441
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 305)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2041  row: L  column: 5
            Class: BAC ends
            High quality sequence stop: 305.
FEATURES             Location/Qualifiers
     source          1. .305
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2041 Col=5 Row=L"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      98 a      54 c      67 g      86 t
```



```
ORIGIN


  Query Match             0.3%;  Score 20;  DB 150;  Length 305;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db     211 AAAAAAAAAAAAATGCTTG 192


RESULT 178
BE954589/c
LOCUS       BE954589      306 bp     mRNA            EST       04-OCT-2000
DEFINITION  UI-M-CC1-bag-e-11-0-UI.s1 NIH_BMAP_Ret1_N Mus musculus cDNA clone
            UI-M-CC1-bag-e-11-0-UI 3', mRNA sequence.
ACCESSION   BE954589
VERSION     BE954589.1  GI:10597275
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail.  The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            embryonic retina tissue cDNA Library Preparation; M.B. Soares Lab
            Clone distribution: Researchers may obtain BMAP cDNA clones from
            RESEARCH GENETICS. It should be noted that Bento Soares is
            generating a small number of additional specialized non-redundant
            arrays of BMAP cDNAs whose availability will be considered under
            appropriate and limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
```



```
     source          1. .306
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-CC1-bag-e-11-0-UI"
                     /clone_lib="NIH_BMAP_Ret1_N"
                     /dev_stage="13.5 days pc"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret1_N library is a normalized library derived
                     NIH_BMAP_Ret1. NIH_BMAP_Ret1 was made from  mouse
                     embryonic retina tissue. For a detailed description of the
                     library from which this clone was  derived, please visit
                     our web site at brainest.eng.uiowa.edu.
                     TAG_LIB=NIH_BMAP_Ret1_N
                     TAG_TISSUE=embryonic-retina
                     TAG_SEQ=CTGCTAGTANCA"
BASE COUNT      92 a      41 c      65 g     108 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 138;  Length 306;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      22 TTAATGAAAAAAAAAAAAAA 3


RESULT 179
AW028947/c
LOCUS       AW028947      310 bp     mRNA            EST       15-SEP-1999
DEFINITION  wv91h10.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2541571 3',
            mRNA sequence.
ACCESSION   AW028947
VERSION     AW028947.1  GI:5887703
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
```

```
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .310
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2541571"
                      /clone_lib="NCI_CGAP_Gas4"
                      /tissue_type="poorly differentiated adenocarcinoma with
                      signet ring cell features"
                      /lab_host="DH10B"
                      /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.69 kb. Life Technologies catalog #:
                      11549-011"
BASE COUNT       92 a     45 c     40 g    133 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 39;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     33 TTAATGAAAAAAAAAAAAAA 14


RESULT 180
BF012355/c
LOCUS       BF012355      310 bp    mRNA            EST       06-OCT-2000
DEFINITION  ux55b02.y1 Soares_NKWMD_mandible Mus musculus cDNA clone
            IMAGE:3514155 5', mRNA sequence.
ACCESSION   BF012355
VERSION     BF012355.1  GI:10712630
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 310)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
```



```
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: ux55b02.x1
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1397035
            Seq primer: -40RP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .310
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3514155"
                      /clone_lib="Soares_NKWMD_mandible"
                      /tissue_type="mandible"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTAATTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       85 a     57 c     54 g    114 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 138;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     45 TTAATGAAAAAAAAAAAAAA 26


RESULT 181
BE995963
LOCUS       BE995963      311 bp    mRNA            EST       05-OCT-2000
DEFINITION  UI-M-BZ1-biw-e-12-0-UI.s1 NIH_BMAP_MHI2_S1 Mus musculus cDNA clone
            UI-M-BZ1-biw-e-12-0-UI 3', mRNA sequence.
ACCESSION   BE995963
VERSION     BE995963.1  GI:10680739
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```



```
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bento Soares
            is generating a small number of additional specialized
            non-redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            The following repetitive elements were found in this cDNA sequence:
            68-89, >AT_rich#Low_complexity 250-296, >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .311
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-BZ1-biw-e-12-0-UI"
                      /clone_lib="NIH_BMAP_MHI2_S1"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_MHI2_S1 library is a subtracted library derived
                      from  NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
                      from mouse hippocampus tissue. For a detailed description
                      of the library from which this clone was  derived, please
                      visit our web site at brainest.eng.uiowa.edu.
                      TAG_SEQ=None found"
BASE COUNT      103 a     63 c     44 g    101 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 138;  Length 311;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
```



```
Db    271 TTAATGAAAAAAAAAAAAAA 290


RESULT 182
AI073068/c
LOCUS       AI073068      312 bp    mRNA            EST       11-FEB-1999
DEFINITION  UI-R-Y0-lx-f-02-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
            UI-R-Y0-lx-f-02-0-UI 3', mRNA sequence.
ACCESSION   AI073068
VERSION     AI073068.1  GI:3399262
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult Eye library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .312
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-Y0-lx-f-02-0-UI"
                      /clone_lib="UI-R-Y0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
                      library is a subtracted library derived from an
                      individually-tagged normalized whole-eye (minus the lens)
                      library. The driver for the subtraction consisted of a
                      pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
                      UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
```

3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
BASE COUNT 77 a 70 c 66 g 99 t
ORIGIN

Query Match 0.3%; Score 20; DB 15; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAAA 4

RESULT 183
AW768811/c
LOCUS AW768811 314 bp mRNA EST 04-MAY-2000
DEFINITION hl53f09.x1 NCI_CGAP_Co14 Homo sapiens cDNA clone IMAGE:3004937 3' similar to gb:X16869 ELONGATION FACTOR 1-ALPHA 1 (HUMAN);, mRNA sequence.
ACCESSION AW768811
VERSION AW768811.1 GI:7700835
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 314)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 238.
FEATURES Location/Qualifiers
source 1. .314
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3004937"
/clone_lib="NCI_CGAP_Co14"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"
BASE COUNT 103 a 54 c 44 g 113 t
ORIGIN

Query Match 0.3%; Score 20; DB 95; Length 314;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 25 TTAATGAAAAAAAAAAAAAA 6

RESULT 184
BB499835
LOCUS BB499835 314 bp mRNA EST 26-JUL-2000
DEFINITION BB499835 RIKEN full-length enriched, 0 day neonate kidney Mus musculus cDNA clone D630023E14 3', mRNA sequence.
ACCESSION BB499835
VERSION BB499835.1 GI:9476381
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 314)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.

, Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .314
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630023E14"
/clone_lib="RIKEN full-length enriched, 0 day neonate kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'

GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverae transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 63 a 83 c 72 g 96 t
ORIGIN

Query Match 0.3%; Score 20; DB 132; Length 314;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 tttgttttccattttgtttc 94
||||||||||||||||||||
Db 258 TTTGTTTTCCATTTTGTTTC 277

RESULT 185
AW265414/c
LOCUS AW265414 315 bp mRNA EST 28-DEC-1999
DEFINITION xq52d12.x1 NCI_CGAP_HN7 Homo sapiens cDNA clone IMAGE:2754263 3' similar to gb:M98447_rna1 PROTEIN-GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE K (HUMAN);, mRNA sequence.
ACCESSION AW265414
VERSION AW265414.1 GI:6642230
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 315)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Edward Shillitoe Ph.D., Silvio Gutkind Ph.D., Chidchanok Leethanakul D.D.S., Michael Emmert-Buck M.D. Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.

```
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2754263"
                     /clone_lib="NCI_CGAP_HN7"
                     /tissue_type="normal squamous epithelium, floor of mouth"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; cDNA made by oligo-dT priming.
                     Non-directionally cloned into the UDG sites of pAMP10.
                     Size-selected on agarose gel, average insert  size 500 bp.
                     Primary library; non-amplified.   cDNA Library
                     Preparation: David B. Krizman, Ph.D (NCI). Reference:
                     Krizman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT       55 a     94 c     73 g     93 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 88;  Length 315;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     42 TTAATGAAAAAAAAAAAAAA 23


RESULT 186
BB276510
LOCUS       BB276510       315 bp    mRNA            EST       07-JUL-2000
DEFINITION  BB276510 RIKEN full-length enriched, 10 days neonate cortex Mus
            musculus cDNA clone A830097M24 3', mRNA sequence.
ACCESSION   BB276510
VERSION     BB276510.1  GI:8973531
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,P., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
```



```
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Owa,C., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A830097M24"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     cortex"
                     /tissue_type="cortex"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGACGTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
```



```
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT      104 a     41 c     52 g    117 t      1 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 126;  Length 315;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    396 aaaaaaatgcttgatattgt 415
          ||||||||||||||||||||
Db    119 AAAAAAATGCTTGATATTGT 138


RESULT 187
AI682836/c
LOCUS       AI682836       318 bp    mRNA            EST       17-DEC-1999
DEFINITION  wc66h12.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2323655 3',
            mRNA sequence.
ACCESSION   AI682836
VERSION     AI682836.1  GI:4893018
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 318)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1753  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 307.
FEATURES             Location/Qualifiers
     source          1. .318
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2323655"
                     /clone_lib="NCI_CGAP_Pan1"
```



```
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT       77 a     65 c     83 g     92 t      1 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 23;  Length 318;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db     27 TTAATGAAAAAAAAAAAAAA 8


RESULT 188
BB021002/c
LOCUS       BB021002       318 bp    mRNA            EST       23-JUN-2000
DEFINITION  BB021002 RIKEN full-length enriched, adult male pituitary gland Mus
            musculus cDNA clone 5330404H06 3', mRNA sequence.
ACCESSION   BB021002
VERSION     BB021002.1  GI:8193195
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 318)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,P., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
```

```
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .318
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5330404H06"
                     /clone_lib="RIKEN full-length enriched, adult male
                     pituitary gland"
                     /sex="male"
                     /tissue_type="pituitary gland"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in House Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 3.0 and subtraction to Rot = 100.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGATTCTCGAGTTAATTAATTAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT       78 a      30 c      43 g     167 t
ORIGIN
```



```
  Query Match             0.3%;  Score 20;  DB 98;  Length 318;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaa 401
           ||||||||||||||||||||
Db     241 ATTAATGAAAAAAAAAAAAA 222


RESULT 189
AI471617/c
LOCUS       AI471617      319 bp    mRNA            EST       09-MAR-1999
DEFINITION  ti66a11.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2136956 3',
            mRNA sequence.
ACCESSION   AI471617
VERSION     AI471617.1  GI:4333707
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 312.
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2136956"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
```



```
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT       74 a      78 c      87 g      80 t
ORIGIN


  Query Match             0.3%;  Score 20;  DB 20;  Length 319;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      20 TTAATGAAAAAAAAAAAAA 1


RESULT 190
AI753085/c
LOCUS       AI753085      319 bp    mRNA            EST       22-JUN-1999
DEFINITION  cr04h02.x2 Jia bone marrow stroma Homo sapiens cDNA clone
            HBMSC_cr04h02 3', mRNA sequence.
ACCESSION   AI753085
VERSION     AI753085.1  GI:5131349
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G.,
            Beckstrom-Sternberg,S.M., Green,E.D., Povell,J.I., Yang,L.M., Robey
            ,P.G., Hotchkiss,R.N. and Francomano,C.A.
  TITLE     SGAP: The Skeletal Genome Anatomy Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Libin Jia
            Medical Genetics Branch
            National Human Genome Research Institute
            10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
            Tel: 301-402-4877
            Fax: 301-496-7157
            Email: libin@helix.nih.gov
            DNA Sequencing and analyses by National Institutes of Health
            Intramural Sequencing Center (NISC).
            Plate: 04  row: h  column: 02
            Seq primer: -21M13 forward primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .319
```



```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HBMSC_cr04h02"
                     /clone_lib="Jia bone marrow stroma"
                     /sex="mixed"
                     /tissue_type="bone marrow stroma"
                     /dev_stage="mixed"
                     /lab_host="XL1-Blue MRF'/SOLR"
                     /note="Vector: pBluescript; Site_1: EcoRI; Site_2: XhoI;
                     mRNA made from human bone marrow stroma, cDNA made by
                     oligo-dT priming. Directionally cloned. Size-selected for
                     average insert size >0.5 kb. Library supplied by Dr. Libin
                     Jia (NHGRI)."
BASE COUNT       83 a      47 c      40 g     146 t       3 others
ORIGIN


  Query Match             0.3%;  Score 20;  DB 24;  Length 319;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db      29 TTAATGAAAAAAAAAAAAA 10


RESULT 191
AQ567887/c
LOCUS       AQ567887      319 bp    DNA             GSS       01-JUN-1999
DEFINITION  HS_2117_A2_D11_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2117 Col=22 Row=G, DNA sequence.
ACCESSION   AQ567887
VERSION     AQ567887.1  GI:4960516
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
```

```
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2117  row: G  column: 22
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 319.
  FEATURES             Location/Qualifiers
       source          1..319
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="Plate=2117 Col=22 Row=G"
                       /clone_lib="CIT Approved Human Genomic Sperm Library D"
                       /sex="male"
                       /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                       E-Coli DH10B"
  BASE COUNT      108 a      49 c      67 g      95 t
  ORIGIN


    Query Match             0.3%;  Score 20;  DB 161;  Length 319;
    Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
    Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy    383 ttaatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||||
  Db     40 TTAATGAAAAAAAAAAAAAA 21


  RESULT 192
  AW457128/c
  LOCUS       AW457128       320 bp    mRNA            EST       24-FEB-2000
  DEFINITION  UI-M-BH3-arc-f-10-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
              UI-M-BH3-arc-f-10-0-UI 3', mRNA sequence.
  ACCESSION   AW457128
  VERSION     AW457128.1  GI:7027345
  KEYWORDS    EST.
  SOURCE      house mouse.
    ORGANISM  Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE   1  (bases 1 to 320)
    AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
    TITLE     Normalization and subtraction: two approaches to facilitate gene
              discovery
    JOURNAL   Genome Res. 6 (9), 791-806 (1996)
    MEDLINE   97044477
  COMMENT     Contact: Chin, H
              National Institute of Mental Health
              6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
```



```
              20892-9643, USA
              Tel: 301 443 1706
              Fax: 301 443 9890
              Email: mEST@mail.nih.gov
              The sequence contained an oligo-dT track that was present in the
              oligonucleotide that was used to prime the synthesis of first
              strand cDNA and therefore this may represent a bonafide poly A
              tail.  The sequence tag present in the cDNA between the NotI site
              and the oligo-dT track served to identify it as a clone from the
              normalized hippocampus library cDNA Library Preparation: M.B.
              Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
              clones from RESEARCH GENETICS. It should be noted that Bento Soares
              is generating a small number of additional specialized
              non-redundant arrays of BMAP cDNAs whose availability will be
              considered under appropriate and limited collaborative arrangements
              Seq primer: M13 Forward
              POLYA=Yes.
  FEATURES             Location/Qualifiers
       source          1..320
                       /organism="Mus musculus"
                       /strain="C57BL/6J"
                       /db_xref="taxon:10090"
                       /clone="UI-M-BH3-arc-f-10-0-UI"
                       /clone_lib="NIH_BMAP_M_S4"
                       /dev_stage="27-32 days"
                       /lab_host="DH10B (Life Technologies)"
                       /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                       polylinker; Site_1: Not I; Site_2: Eco RI; The
                       NIH_BMAP_M_S4 library is a subtracted library of a series,
                       ultimately  derived from a mixture of individually tagged
                       normalized libraries from  ten regions of the mouse brain
                       (cerebellum, brain stems, olfactory  bulbs, hypothalamus,
                       cortex, amygdala, basal ganglia, pineal gland,  striatum,
                       hipoccampus) after a series of subtractions to reduce the
                       representation of cDNAs from which ESTs had already been
                       generated. The following serially subtracted libraries
                       were generated in this process: NIH_BMAP_M_S4,
                       NIH_BMAP_M_S3.3,  NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                       NIH_BMAP_M_S2, NIH_BMAP_M_S1.  The subtracted library
                       (NIH_BMAP_M_S4) was constructed as follows:  PCRamplified
                       cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                       NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                       was used as  a driver in a hybridization with a pool of
                       the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                       libraries in the form of  single-stranded circles.  The
                       remaining single-stranded circles  (subtracted library)
                       was purified by hydroxyapatite column chromatography,
                       converted to double-stranded circles and electroporated
                       into DH10B  bacteria (LifeTechnologies) to generate the
                       NIH_BMAP_M_S4 library.  This procedure has been previously
                       described (Bonaldo, Lennon and Soares, Genome Research
```



```
                       6:791-806, 1996)
                       TAG_LIB=NIH_BMAP_M_S4
                       TAG_TISSUE=hippocampus
                       TAG_SEQ=TTCCA"
  BASE COUNT       58 a      79 c      79 g     104 t
  ORIGIN


    Query Match             0.3%;  Score 20;  DB 90;  Length 320;
    Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
    Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy    383 ttaatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||||
  Db     23 TTAATGAAAAAAAAAAAAAA 4


  RESULT 193
  BE995761/c
  LOCUS       BE995761       320 bp    mRNA            EST       05-OCT-2000
  DEFINITION  UI-M-BZ1-biv-c-04-0-UI.s1 NIH_BMAP_MHI2_S1 Mus musculus cDNA clone
              UI-M-BZ1-biv-c-04-0-UI 3', mRNA sequence.
  ACCESSION   BE995761
  VERSION     BE995761.1  GI:10680337
  KEYWORDS    EST.
  SOURCE      house mouse.
    ORGANISM  Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE   1  (bases 1 to 320)
    AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
    TITLE     Normalization and subtraction: two approaches to facilitate gene
              discovery
    JOURNAL   Genome Res. 6 (9), 791-806 (1996)
    MEDLINE   97044477
  COMMENT     Contact: Chin, H
              National Institute of Mental Health
              6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
              20892-9643, USA
              Tel: 301 443 1706
              Fax: 301 443 9890
              Email: mEST@mail.nih.gov
              The sequence contained an oligo-dT track that was present in the
              oligonucleotide that was used to prime the synthesis of first
              strand cDNA and therefore this may represent a bonafide poly A
              tail. The sequence tag present in the cDNA between the NotI site
              and the oligo-dT track served to verify it as a clone from the
              hippocampus tissue cDNA Library Preparation: M.B. Soares Lab Clone
              distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
              GENETICS. It should be noted that Bento Soares is generating a
              small number of additional specialized non-redundant arrays of BMAP
```



```
              cDNAs whose availability will be considered under appropriate and
              limited collaborative arrangements
              Seq primer: M13 Forward
              POLYA=Yes.
  FEATURES             Location/Qualifiers
       source          1..320
                       /organism="Mus musculus"
                       /strain="C57BL/6J"
                       /db_xref="taxon:10090"
                       /clone="UI-M-BZ1-biv-c-04-0-UI"
                       /clone_lib="NIH_BMAP_MHI2_S1"
                       /dev_stage="27-32 days"
                       /lab_host="DH10B (Life Technologies)"
                       /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                       polylinker; Site_1: Not I; Site_2: Eco RI; The
                       NIH_BMAP_MHI2_S1 library is a subtracted library derived
                       from  NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
                       from mouse hippocampus tissue. For a detailed description
                       of the library from which this clone was  derived, please
                       visit our web site at brainest.eng.uiowa.edu.
                       TAG_LIB=NIH_BMAP_MHI2_S1
                       TAG_TISSUE=hippocampus
                       TAG_SEQ=TAGCC"
  BASE COUNT       88 a      85 c      69 g      78 t
  ORIGIN


    Query Match             0.3%;  Score 20;  DB 138;  Length 320;
    Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
    Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy    383 ttaatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||||
  Db     24 TTAATGAAAAAAAAAAAAAA 5


  RESULT 194
  AQ071286
  LOCUS       AQ071286       321 bp    DNA             GSS       05-AUG-1998
  DEFINITION  HS_2266_B2_C06_MR CIT Approved Human Genomic Sperm Library D Homo
              sapiens genomic clone Plate=2266 Col=12 Row=F, DNA sequence.
  ACCESSION   AQ071286
  VERSION     AQ071286.1  GI:3388445
  KEYWORDS    GSS.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 321)
    AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
              Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
```

```
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2266  row: F  column: 12
            Class: BAC ends
            High quality sequence stop: 321.
FEATURES              Location/Qualifiers
     source           1. .321
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=2266 Col=12 Row=F"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      100 a     80 c     69 g     72 t
ORIGIN


  Query Match           0.3%;  Score 20;  DB 148;  Length 321;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3218 aagttaataaaaagaaagca 3237
          ||||||||||||||||||||
Db    170 AAGTTAATAAAAAGAAAGCA 189


RESULT 195
AI248370
LOCUS       AI248370      324 bp    mRNA            EST       01-DEC-1998
DEFINITION  qh76b01.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:1850569 3', mRNA sequence.
ACCESSION   AI248370
VERSION     AI248370.1  GI:3843767
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 324)
```



```
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 935    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 280.
FEATURES              Location/Qualifiers
     source           1. .324
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1850569"
                      /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      This is a subtracted version of the original Soares fetal
                      liver spleen 1NFLS library.   1st strand cDNA was primed
                      with a Pac I - oligo(dT) primer [5'
                      AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.   Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      103 a     68 c     76 g     77 t
ORIGIN


  Query Match           0.3%;  Score 20;  DB 18;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db    108 TTAATGAAAAAAAAAAAAAA 127


RESULT 196
AW619259
LOCUS       AW619259      324 bp    mRNA            EST       24-MAR-2000
DEFINITION  619 MARC PBE Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   AW619259
VERSION     AW619259.1  GI:7325443
```



```
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Smith,T.P.L., Fahrenkrug,S.C., Rohrer,G.A., Simmen,F.A., Rexroad
            ,C.E. and Keele,J.W.
  TITLE     Mapping of Expressed Sequence Tags from a porcine early embryonic
            cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: GGAAACAGCTATGACCATG
            BACKWARD: GTAAAACGACGGCCAGT
            Seq primer: AATTAACCCTCACTAAAGGG.
FEATURES              Location/Qualifiers
     source           1. .324
                      /organism="Sus scrofa"
                      /db_xref="taxon:9823"
                      /clone_lib="MARC PBE"
                      /tissue_type="Day 12 whole embryos"
                      /lab_host="XLOLR"
                      /note="Vector: pBLUESCRIPT SK-; Site_1: EcoRI; Site_2:
                      XhoI; Library made from pool of embryos in spherical and
                      filamentous stages of development (7.5% and 92.5%,
                      respectively, of each stage) as described in Choi et al,
                      Endocrinology 137, 1457-67, 1996."
BASE COUNT      109 a     57 c     65 g     93 t
ORIGIN


  Query Match           0.3%;  Score 20;  DB 93;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    383 ttaatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||||
Db    302 TTAATGAAAAAAAAAAAAAA 321


RESULT 197
AW528408/c
```



```
LOCUS       AW528408      327 bp    mRNA            EST       06-MAR-2000
DEFINITION  UI-R-BO1-ajl-b-01-0-UI.s1 UI-R-BO1 Rattus norvegicus cDNA clone
            UI-R-BO1-ajl-b-01-0-UI 3', mRNA sequence.
ACCESSION   AW528408
VERSION     AW528408.1  GI:7170822
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized corpus-striatum library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .327
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-BO1-ajl-b-01-0-UI"
                      /clone_lib="UI-R-BO1"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The library
                      (UI-R-BO1) is a subtracted library derived from a mixture
                      of the following tissues: thalamus, cerebellum,
                      hypothalamus, medulla, pons, midbrain, cerebral cortex,
                      corpus striatum and hippocampus. For a detailed
                      description of the library from which this clone was
                      derived, please visit our web site at
```

ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-B01
TAG_TISSUE=corpus-striatum
TAG_SEQ="CTAGG"
BASE COUNT 70 a 85 c 84 g 88 t
ORIGIN

Query Match 0.3%; Score 20; DB 91; Length 327;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAAA 4

RESULT 198
AI070265/c
LOCUS AI070265 328 bp mRNA EST 05-JUL-1999
DEFINITION UI-R-Y0-ls-f-10-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone UI-R-Y0-ls-f-10-0-UI 3', mRNA sequence.
ACCESSION AI070265
VERSION AI070265.1 GI:3396516
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 328)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 9250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Eye library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1785279 The

following repetitive elements were found in this cDNA sequence:
1-43, >AT_rich#Low_complexity
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. .328
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-ls-f-10-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually-tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
BASE COUNT 85 a 66 c 53 g 124 t
ORIGIN

Query Match 0.3%; Score 20; DB 15; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 23 TTAATGAAAAAAAAAAAAAA 4

RESULT 199
AL047858
LOCUS AL047858 330 bp mRNA EST 01-MAR-2000

DEFINITION DKFZp586L1322_r1 586 (synonym: hutel) Homo sapiens cDNA clone DKFZp586L1322, mRNA sequence.
ACCESSION AL047858
VERSION AL047858.1 GI:4728046
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 330)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Koehrer, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Koehrer K
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ): Email s.wiemann@dkfz- heidelberg.de; sequenced by BMPZ (Biomedical Research Center at the Charite, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
No a1 sequence available.
This clone (DKFZp586L1322) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES Location/Qualifiers
source 1. .330
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp586L1322"
/clone_lib="586 (synonym: hutel)"
/tissue_type="uterus"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSportl; Site_1: NotI; Site_2: SalI/MluI"
BASE COUNT 84 a 91 c 72 g 83 t
ORIGIN

Query Match 0.3%; Score 20; DB 28; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 ttaatgaaaaaaaaaaaaaa 402
||||||||||||||||||||
Db 307 TTAATGAAAAAAAAAAAAAA 326

RESULT 200
H42282

LOCUS H42282 330 bp mRNA EST 31-JUL-1995
DEFINITION yo63a03.r1 Soares breast 3NbHBst Homo sapiens cDNA clone IMAGE:182597 5', mRNA sequence.
ACCESSION H42282
VERSION H42282.1 GI:918334
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 330)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1802
High quality sequence stops: 263
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1802 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 263.
FEATURES Location/Qualifiers
source 1. .330
/organism="Homo sapiens"
/db_xref="GDB:3815797"
/db_xref="taxon:9606"
/clone="IMAGE:182597"
/clone_lib="Soares breast 3NbHBst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 117 a 36 c 64 g 113 t
ORIGIN

Query Match 0.3%; Score 20; DB 141; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    5026 tttgacttttaaatttttatt 5045
           ||||||||||||||||||||
Db     146 TTTGACTTTAAATTTTTATT 165
```

Search completed: March 26, 2001, 17:53:03
Job time: 20232 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 14:20:13 ; Search time 180.84 Seconds
(without alignments)
5280.219 Million cell updates/sec

Title: US-09-410-835-4
Perfect score: 5925
Sequence: 1 ctccattttaagcaactcca.....ttttagaatagacaaagctt 5925

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 280836 seqs, 80580151 residues

Word size: 15

Total number of hits satisfying chosen parameters: 3180

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_NA:*
1: /cgn1_7/ptodata/1/ina/5A_COMB.seq:*
2: /cgn1_7/ptodata/1/ina/5B_COMB.seq:*
3: /cgn1_7/ptodata/1/ina/6_COMB.seq:*
4: /cgn1_7/ptodata/1/ina/PCTUS_COMB.seq:*
5: /cgn1_7/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5925	100.0	5925	3	US-08-470-260-4	Sequence 4, Appli
2	5925	100.0	5925	3	US-08-471-491-4	Sequence 4, Appli
3	5925	100.0	5925	3	US-08-466-662-4	Sequence 4, Appli
4	5925	100.0	19932	2	US-08-477-451-25	Sequence 25, Appl
5	5874	99.1	10299	2	US-08-477-451-1	Sequence 1, Appli
c 6	5772	97.4	10299	2	US-08-477-451-5	Sequence 5, Appli

7	235	4.0	3648	1	US-08-053-614-1	Sequence 1, Appli
8	235	4.0	3648	1	US-08-316-397B-1	Sequence 1, Appli
9	235	4.0	3648	2	US-09-034-306-1	Sequence 1, Appli
10	235	4.0	3648	3	US-09-259-437-1	Sequence 1, Appli
11	235	4.0	3648	4	PCT-US93-09782-1	Sequence 1, Appli
12	235	4.0	4821	1	US-08-053-614-3	Sequence 3, Appli
13	235	4.0	4821	1	US-08-316-397B-3	Sequence 3, Appli
14	235	4.0	4821	2	US-09-034-306-3	Sequence 3, Appli
15	235	4.0	4821	3	US-09-259-437-3	Sequence 3, Appli
16	235	4.0	4821	4	PCT-US93-09782-3	Sequence 3, Appli
17	21	0.4	1604	3	US-08-965-903B-5	Sequence 5, Appli
c 18	20	0.3	20	3	US-09-012-198-6	Sequence 6, Appli
19	20	0.3	284	1	US-08-377-687-50	Sequence 50, Appl
20	20	0.3	284	1	US-08-777-192-50	Sequence 50, Appl
21	20	0.3	1537	2	US-08-839-008-1	Sequence 1, Appli
22	20	0.3	2173	1	US-08-036-210-6	Sequence 6, Appli
23	20	0.3	2173	2	US-08-449-609-6	Sequence 6, Appli
24	20	0.3	2309	1	US-08-036-210-10	Sequence 10, Appl
25	20	0.3	2309	2	US-08-449-609-10	Sequence 10, Appl
c 26	20	0.3	2529	4	PCT-US91-04541-1	Sequence 1, Appli
c 27	20	0.3	2529	5	5262319-1	Patent No. 5262319
28	20	0.3	2692	1	US-08-036-210-14	Sequence 14, Appl
29	20	0.3	2692	2	US-08-449-609-14	Sequence 14, Appl
30	20	0.3	3973	1	US-08-036-210-21	Sequence 21, Appl
31	20	0.3	3973	2	US-08-449-609-21	Sequence 21, Appl
c 32	19	0.3	35	1	US-08-126-594-13	Sequence 13, Appl
c 33	19	0.3	35	1	US-08-465-811A-13	Sequence 13, Appl
c 34	19	0.3	35	2	US-08-619-542B-13	Sequence 13, Appl
35	19	0.3	4089	1	US-07-908-245-1	Sequence 1, Appli
c 36	19	0.3	4748	3	US-09-331-581-1	Sequence 1, Appli
c 37	19	0.3	7286	3	US-09-331-581-3	Sequence 3, Appli
c 38	19	0.3	7938	3	US-09-331-581-14	Sequence 14, Appl
c 39	18	0.3	37	2	US-08-870-010A-1	Sequence 1, Appli
c 40	18	0.3	37	3	US-09-102-371-1	Sequence 1, Appli
41	18	0.3	39	2	US-08-975-902-47	Sequence 47, Appl
42	18	0.3	39	3	US-09-251-565-47	Sequence 47, Appl
c 43	18	0.3	40	1	US-08-510-032A-3	Sequence 3, Appli
c 44	18	0.3	40	3	US-08-688-514-3	Sequence 3, Appli
45	18	0.3	309	1	US-08-086-410-24	Sequence 24, Appl
46	18	0.3	360	1	US-07-920-519-28	Sequence 28, Appl
47	18	0.3	360	1	US-08-086-410-21	Sequence 21, Appl
48	18	0.3	360	1	US-08-314-586-28	Sequence 28, Appl
49	18	0.3	507	1	US-08-686-878A-22	Sequence 22, Appl
50	18	0.3	861	2	US-08-924-759-15	Sequence 15, Appl
51	18	0.3	861	3	US-09-248-335-15	Sequence 15, Appl
52	18	0.3	882	2	US-08-909-965C-9	Sequence 9, Appli
53	18	0.3	940	1	US-08-089-998B-1	Sequence 1, Appli
54	18	0.3	940	1	US-08-457-272-1	Sequence 1, Appli
55	18	0.3	940	4	PCT-US94-07595-1	Sequence 1, Appli
56	18	0.3	1250	3	US-08-688-988-1	Sequence 1, Appli
57	18	0.3	1314	3	US-09-025-059-2	Sequence 2, Appli

c 58	18	0.3	1357	1	US-08-270-076A-1	Sequence 1, Appli
59	18	0.3	1373	2	US-08-394-189B-3	Sequence 3, Appli
60	18	0.3	1373	4	PCT-US93-05705-3	Sequence 3, Appli
c 61	18	0.3	1483	1	US-08-270-076A-2	Sequence 2, Appli
62	18	0.3	1539	3	US-08-582-740-69	Sequence 69, Appl
63	18	0.3	1573	1	US-08-597-545-4	Sequence 4, Appli
64	18	0.3	1573	1	US-08-457-135-4	Sequence 4, Appli
65	18	0.3	1659	1	US-08-203-716-1	Sequence 1, Appli
66	18	0.3	1659	1	US-08-440-579-1	Sequence 1, Appli
67	18	0.3	1659	3	US-09-039-657-1	Sequence 1, Appli
68	18	0.3	1659	4	PCT-US91-02339-2	Sequence 2, Appli
69	18	0.3	1694	1	US-08-698-551-9	Sequence 9, Appli
70	18	0.3	1694	2	US-08-602-228-9	Sequence 9, Appli
71	18	0.3	1694	2	US-08-494-440B-9	Sequence 9, Appli
72	18	0.3	1694	2	US-08-533-901B-9	Sequence 9, Appli
73	18	0.3	1694	2	US-08-839-032A-9	Sequence 9, Appli
74	18	0.3	1694	2	US-08-839-031A-9	Sequence 9, Appli
75	18	0.3	1694	4	PCT-US95-12724-9	Sequence 9, Appli
76	18	0.3	1822	1	US-07-841-646-1	Sequence 1, Appli
77	18	0.3	1822	1	US-07-901-703-1	Sequence 1, Appli
78	18	0.3	1822	1	US-08-147-023-1	Sequence 1, Appli
79	18	0.3	1822	1	US-08-206-864-1	Sequence 1, Appli
80	18	0.3	1822	1	US-08-278-729A-16	Sequence 16, Appl
81	18	0.3	1822	1	US-08-480-528A-3	Sequence 3, Appli
82	18	0.3	1822	1	US-08-479-666-3	Sequence 3, Appli
83	18	0.3	1822	1	US-09-155-343A-16	Sequence 16, Appl
84	18	0.3	1822	1	US-08-406-672-16	Sequence 16, Appl
85	18	0.3	1822	1	US-08-643-563A-16	Sequence 16, Appl
86	18	0.3	1822	1	US-08-447-570-1	Sequence 1, Appli
87	18	0.3	1822	1	US-08-643-763A-16	Sequence 16, Appl
88	18	0.3	1822	1	US-08-462-623-16	Sequence 16, Appl
89	18	0.3	1822	1	US-08-451-953A-16	Sequence 16, Appl
90	18	0.3	1822	2	US-08-459-346-1	Sequence 1, Appli
91	18	0.3	1822	2	US-08-445-468A-16	Sequence 16, Appl
92	18	0.3	1822	2	US-08-901-200A-3	Sequence 3, Appli
93	18	0.3	1822	2	US-08-481-337A-9	Sequence 9, Appli
94	18	0.3	1822	2	US-08-449-700-1	Sequence 1, Appli
95	18	0.3	1822	2	US-08-449-699A-1	Sequence 1, Appli
96	18	0.3	1822	2	US-08-696-268B-3	Sequence 3, Appli
97	18	0.3	1822	2	US-08-461-397A-16	Sequence 16, Appl
98	18	0.3	1822	2	US-08-912-088-16	Sequence 16, Appl
99	18	0.3	1822	3	US-08-278-730A-16	Sequence 16, Appl
100	18	0.3	1822	3	US-08-458-811-1	Sequence 1, Appli
101	18	0.3	1822	3	US-08-889-419-1	Sequence 1, Appli
102	18	0.3	1822	3	US-08-445-467-16	Sequence 16, Appl
103	18	0.3	1822	3	US-08-480-515A-16	Sequence 16, Appl
104	18	0.3	1822	3	US-08-459-129-1	Sequence 1, Appli
105	18	0.3	1822	3	US-09-219-391-3	Sequence 3, Appli
106	18	0.3	1822	4	PCT-US90-07654-3	Sequence 3, Appli
107	18	0.3	1822	4	PCT-US91-07635-1	Sequence 1, Appli
108	18	0.3	1822	4	PCT-US92-01968-16	Sequence 16, Appl

109	18	0.3	1822	4	PCT-US93-05446-1	Sequence 1, Appli
110	18	0.3	1822	4	PCT-US93-07189-1	Sequence 1, Appli
111	18	0.3	1822	4	PCT-US93-07190-16	Sequence 16, Appl
112	18	0.3	1822	4	PCT-US93-07231-16	Sequence 16, Appl
113	18	0.3	1822	4	PCT-US93-08742-16	Sequence 16, Appl
114	18	0.3	1822	4	PCT-US93-08808-16	Sequence 16, Appl
115	18	0.3	1822	4	PCT-US93-08885-16	Sequence 16, Appl
116	18	0.3	1822	4	PCT-US93-10520-3	Sequence 3, Appli
117	18	0.3	1822	4	PCT-US95-05467-9	Sequence 9, Appli
118	18	0.3	1822	4	PCT-US95-06724-1	Sequence 1, Appli
119	18	0.3	1855	3	US-09-150-133-6	Sequence 6, Appli
120	18	0.3	1855	3	US-09-150-141-6	Sequence 6, Appli
121	18	0.3	1873	1	US-07-841-646-24	Sequence 24, Appl
122	18	0.3	1873	1	US-07-901-703-8	Sequence 8, Appli
123	18	0.3	1873	1	US-08-147-023-24	Sequence 24, Appl
124	18	0.3	1873	1	US-08-278-729A-18	Sequence 18, Appl
125	18	0.3	1873	1	US-08-480-528A-5	Sequence 5, Appli
126	18	0.3	1873	1	US-08-479-666-5	Sequence 5, Appli
127	18	0.3	1873	1	US-08-155-343A-18	Sequence 18, Appl
128	18	0.3	1873	1	US-08-406-672-18	Sequence 18, Appl
129	18	0.3	1873	1	US-08-643-563A-18	Sequence 18, Appl
130	18	0.3	1873	1	US-08-447-570-24	Sequence 24, Appl
131	18	0.3	1873	1	US-08-643-763A-18	Sequence 18, Appl
132	18	0.3	1873	1	US-08-462-623-18	Sequence 18, Appl
133	18	0.3	1873	1	US-08-451-953A-18	Sequence 18, Appl
134	18	0.3	1873	2	US-08-459-346-3	Sequence 3, Appli
135	18	0.3	1873	2	US-08-445-468A-18	Sequence 18, Appl
136	18	0.3	1873	2	US-08-901-200A-5	Sequence 5, Appli
137	18	0.3	1873	2	US-08-449-700-24	Sequence 24, Appl
138	18	0.3	1873	2	US-08-449-699A-24	Sequence 24, Appl
139	18	0.3	1873	2	US-08-461-397A-18	Sequence 18, Appl
140	18	0.3	1873	2	US-08-912-088-18	Sequence 18, Appl
141	18	0.3	1873	3	US-08-278-730A-18	Sequence 18, Appl
142	18	0.3	1873	3	US-08-889-419-3	Sequence 3, Appli
143	18	0.3	1873	3	US-08-445-467-18	Sequence 18, Appl
144	18	0.3	1873	3	US-08-480-515A-18	Sequence 18, Appl
145	18	0.3	1873	3	US-09-219-391-5	Sequence 5, Appli
146	18	0.3	1873	4	PCT-US90-07654-1	Sequence 1, Appli
147	18	0.3	1873	4	PCT-US92-01968-18	Sequence 18, Appl
148	18	0.3	1873	4	PCT-US93-05446-8	Sequence 8, Appli
149	18	0.3	1873	4	PCT-US93-07189-3	Sequence 3, Appli
150	18	0.3	1873	4	PCT-US93-07190-18	Sequence 18, Appl
151	18	0.3	1873	4	PCT-US93-07231-18	Sequence 18, Appl
152	18	0.3	1873	4	PCT-US93-08742-18	Sequence 18, Appl
153	18	0.3	1873	4	PCT-US93-08808-18	Sequence 18, Appl
154	18	0.3	1873	4	PCT-US93-08885-18	Sequence 18, Appl
155	18	0.3	1873	4	PCT-US93-10520-5	Sequence 5, Appli
156	18	0.3	1882	1	US-08-458-120-1	Sequence 1, Appli
157	18	0.3	1882	2	US-08-867-970-1	Sequence 1, Appli
158	18	0.3	1905	3	US-08-387-117-7	Sequence 7, Appli
159	18	0.3	1982	2	US-08-900-148-1	Sequence 1, Appli

	160	18	0.3	2164	1	US-08-096-277-6 Sequence 6, Appli
	161	18	0.3	2164	2	US-08-550-815-6 Sequence 6, Appli
	162	18	0.3	2164	3	US-08-703-089-6 Sequence 6, Appli
	163	18	0.3	2199	1	US-07-847-743B-7 Sequence 7, Appli
	164	18	0.3	2199	1	US-08-096-277-7 Sequence 7, Appli
	165	18	0.3	2199	1	US-08-456-201-7 Sequence 7, Appli
	166	18	0.3	2199	2	US-08-456-241-7 Sequence 7, Appli
	167	18	0.3	2199	2	US-08-550-815-7 Sequence 7, Appli
	168	18	0.3	2199	3	US-08-703-089-7 Sequence 7, Appli
	169	18	0.3	2199	4	PCT-US92-04295A-7 Sequence 7, Appli
	170	18	0.3	2524	1	US-08-317-522A-1 Sequence 1, Appli
	171	18	0.3	2524	1	US-08-439-818A-1 Sequence 1, Appli
	172	18	0.3	2524	2	US-08-751-965-1 Sequence 1, Appli
	173	18	0.3	2524	2	US-08-738-975-1 Sequence 1, Appli
	174	18	0.3	2524	2	US-08-728-626-1 Sequence 1, Appli
	175	18	0.3	2524	3	US-08-808-599A-1 Sequence 1, Appli
	176	18	0.3	2582	1	US-08-514-014-3 Sequence 3, Appli
	177	18	0.3	2582	2	US-08-833-823-3 Sequence 3, Appli
	178	18	0.3	2781	1	US-08-178-477B-42 Sequence 42, Appl
	179	18	0.3	2801	3	US-08-747-221B-30 Sequence 30, Appl
c	180	18	0.3	2801	3	US-08-747-221B-32 Sequence 32, Appl
	181	18	0.3	2836	3	US-08-747-221B-24 Sequence 24, Appl
c	182	18	0.3	2836	3	US-08-747-221B-26 Sequence 26, Appl
	183	18	0.3	2999	2	US-09-014-969-8 Sequence 8, Appli
	184	18	0.3	3349	2	US-08-920-234-1 Sequence 1, Appli
	185	18	0.3	3442	1	US-08-529-654-3 Sequence 3, Appli
c	186	18	0.3	3622	2	US-08-591-079-7 Sequence 7, Appli
	187	18	0.3	4304	4	PCT-US96-00419-1 Sequence 1, Appli
c	188	18	0.3	5393	2	US-08-591-079-9 Sequence 9, Appli
c	189	18	0.3	8491	2	US-08-757-439-1 Sequence 1, Appli
	190	18	0.3	246240	2	US-08-724-394A-20 Sequence 20, Appl
	191	18	0.3	246240	2	US-08-724-394A-21 Sequence 21, Appl
	192	18	0.3	246240	2	US-08-724-394A-22 Sequence 22, Appl
	193	17	0.3	27	2	US-08-553-367A-17 Sequence 17, Appl
c	194	17	0.3	32	2	US-08-132-990A-25 Sequence 25, Appl
c	195	17	0.3	33	3	US-08-961-083-330 Sequence 330, App
c	196	17	0.3	54	2	US-08-418-848A-57 Sequence 57, Appl
	197	17	0.3	85	1	US-08-120-827-97 Sequence 97, Appl
	198	17	0.3	85	1	US-08-478-675-97 Sequence 97, Appl
	199	17	0.3	105	2	US-08-735-381-2 Sequence 2, Appli
	200	17	0.3	105	3	US-09-183-619-1 Sequence 1, Appli
	201	17	0.3	105	3	US-09-201-674-2 Sequence 2, Appli
c	202	17	0.3	123	5	5198348-4 Patent No. 5198348
	203	17	0.3	422	3	US-08-621-018B-47 Sequence 47, Appl
	204	17	0.3	461	2	US-08-467-046-24 Sequence 24, Appl
c	205	17	0.3	464	1	US-08-196-178-2 Sequence 2, Appli
c	206	17	0.3	666	1	US-08-196-178-3 Sequence 3, Appli
	207	17	0.3	739	1	US-08-156-383-3 Sequence 3, Appli
	208	17	0.3	739	1	US-08-340-136-3 Sequence 3, Appli
	209	17	0.3	739	4	PCT-US92-10866-3 Sequence 3, Appli
	210	17	0.3	748	1	US-08-361-467B-3 Sequence 3, Appli

	211	17	0.3	748	1	US-08-484-332C-3 Sequence 3, Appli
	212	17	0.3	775	3	US-09-361-434-6 Sequence 6, Appli
c	213	17	0.3	775	3	US-09-361-434-8 Sequence 8, Appli
c	214	17	0.3	797	2	US-08-752-132-1 Sequence 1, Appli
	215	17	0.3	836	2	US-08-698-805-7 Sequence 7, Appli
	216	17	0.3	894	1	US-08-178-708-7 Sequence 7, Appli
	217	17	0.3	894	1	US-08-457-552-7 Sequence 7, Appli
	218	17	0.3	894	1	US-08-456-430-7 Sequence 7, Appli
	219	17	0.3	894	2	US-08-994-418-7 Sequence 7, Appli
	220	17	0.3	894	4	PCT-US95-00432-7 Sequence 7, Appli
	221	17	0.3	900	1	US-08-336-031-3 Sequence 3, Appli
	222	17	0.3	900	4	PCT-US95-06725-3 Sequence 3, Appli
c	223	17	0.3	908	2	US-08-286-819A-22 Sequence 22, Appl
c	224	17	0.3	909	3	US-08-980-357-22 Sequence 22, Appl
	225	17	0.3	1046	1	US-08-361-467B-4 Sequence 4, Appli
	226	17	0.3	1046	1	US-08-484-332C-4 Sequence 4, Appli
	227	17	0.3	1106	3	US-09-361-434-16 Sequence 16, Appl
c	228	17	0.3	1106	3	US-09-361-434-18 Sequence 18, Appl
	229	17	0.3	1131	2	US-08-951-924A-1 Sequence 1, Appli
	230	17	0.3	1138	3	US-08-946-914-1 Sequence 1, Appli
	231	17	0.3	1147	1	US-08-665-716-1 Sequence 1, Appli
c	232	17	0.3	1211	1	US-08-196-178-1 Sequence 1, Appli
c	233	17	0.3	1222	1	US-08-314-596-43 Sequence 43, Appl
c	234	17	0.3	1222	1	US-08-320-982-43 Sequence 43, Appl
c	235	17	0.3	1222	3	US-08-819-037-43 Sequence 43, Appl
	236	17	0.3	1326	2	US-08-484-993B-5 Sequence 5, Appli
	237	17	0.3	1326	2	US-08-484-158B-5 Sequence 5, Appli
	238	17	0.3	1326	2	US-08-484-596A-5 Sequence 5, Appli
	239	17	0.3	1326	2	US-08-480-150A-5 Sequence 5, Appli
	240	17	0.3	1326	3	US-08-458-731-5 Sequence 5, Appli
	241	17	0.3	1326	3	US-08-149-223A-5 Sequence 5, Appli
c	242	17	0.3	1382	4	PCT-US92-07982A-1 Sequence 1, Appli
	243	17	0.3	1426	3	US-08-705-771-5 Sequence 5, Appli
	244	17	0.3	1430	1	US-08-276-452A-25 Sequence 25, Appl
	245	17	0.3	1430	2	US-08-798-744-25 Sequence 25, Appl
c	246	17	0.3	1448	1	US-08-314-596-39 Sequence 39, Appl
c	247	17	0.3	1448	1	US-08-320-982-39 Sequence 39, Appl
c	248	17	0.3	1448	3	US-08-819-037-39 Sequence 39, Appl
	249	17	0.3	1454	2	US-08-713-000-7 Sequence 7, Appli
	250	17	0.3	1454	2	US-08-975-316-7 Sequence 7, Appli
	251	17	0.3	1474	2	US-08-975-316-71 Sequence 71, Appl
	252	17	0.3	1511	2	US-08-809-763-3 Sequence 3, Appli
	253	17	0.3	1511	3	US-08-956-253-3 Sequence 3, Appli
c	254	17	0.3	1543	2	US-08-991-946A-4 Sequence 4, Appli
	255	17	0.3	1562	1	US-08-330-108-1 Sequence 1, Appli
	256	17	0.3	1562	4	PCT-US92-10087-1 Sequence 1, Appli
	257	17	0.3	1620	3	US-08-985-950-11 Sequence 11, Appl
	258	17	0.3	1636	5	5447867-2 Patent No. 5447867
	259	17	0.3	1641	2	US-09-080-897-5 Sequence 5, Appli
	260	17	0.3	1654	3	US-08-913-842-20 Sequence 20, Appl
	261	17	0.3	1656	3	US-09-072-384-14 Sequence 14, Appl

	262	17	0.3	1679	3	US-09-072-384-17 Sequence 17, Appl
	263	17	0.3	1680	5	5447867-3 Patent No. 5447867
	264	17	0.3	1684	2	US-08-831-570-1 Sequence 1, Appli
	265	17	0.3	1684	2	US-08-831-575-1 Sequence 1, Appli
	266	17	0.3	1685	1	US-08-366-779-4 Sequence 4, Appli
	267	17	0.3	1685	1	US-08-789-936-4 Sequence 4, Appli
	268	17	0.3	1690	1	US-08-276-452A-24 Sequence 24, Appl
	269	17	0.3	1690	2	US-08-798-744-24 Sequence 24, Appl
	270	17	0.3	1703	2	US-08-794-795-1 Sequence 1, Appli
	271	17	0.3	1744	3	US-08-913-842-27 Sequence 27, Appl
	272	17	0.3	1764	2	US-08-504-459-3 Sequence 3, Appli
	273	17	0.3	1764	3	US-09-013-881-13 Sequence 13, Appl
	274	17	0.3	1789	3	US-09-035-706-1 Sequence 1, Appli
	275	17	0.3	1789	3	US-08-955-841-1 Sequence 1, Appli
	276	17	0.3	1856	2	US-08-360-606B-29 Sequence 29, Appl
	277	17	0.3	1875	3	US-08-913-842-1 Sequence 1, Appli
	278	17	0.3	1965	2	US-08-809-763-5 Sequence 5, Appli
	279	17	0.3	1965	3	US-08-913-842-2 Sequence 2, Appli
	280	17	0.3	1965	3	US-08-956-253-5 Sequence 5, Appli
	281	17	0.3	2134	1	US-08-336-031-1 Sequence 1, Appli
	282	17	0.3	2134	4	PCT-US95-06725-1 Sequence 1, Appli
	283	17	0.3	2306	1	US-08-484-105-9 Sequence 9, Appli
	284	17	0.3	2306	1	US-08-484-106-9 Sequence 9, Appli
	285	17	0.3	2360	1	US-08-466-390-1 Sequence 1, Appli
	286	17	0.3	2360	1	US-08-470-950-1 Sequence 1, Appli
	287	17	0.3	2360	1	US-08-467-781-1 Sequence 1, Appli
	288	17	0.3	2360	1	US-08-195-467-1 Sequence 1, Appli
	289	17	0.3	2360	2	US-08-485-924-1 Sequence 1, Appli
	290	17	0.3	2360	3	US-08-836-567-9 Sequence 9, Appli
	291	17	0.3	2360	4	PCT-US93-06160-1 Sequence 1, Appli
	292	17	0.3	2383	1	US-08-274-318-1 Sequence 1, Appli
	293	17	0.3	2383	2	US-08-754-108-1 Sequence 1, Appli
	294	17	0.3	2397	5	5258288-2 Patent No. 5258288
	295	17	0.3	2406	2	US-08-463-081B-1 Sequence 1, Appli
	296	17	0.3	2406	2	US-08-461-379A-1 Sequence 1, Appli
	297	17	0.3	2406	2	US-08-462-390B-1 Sequence 1, Appli
	298	17	0.3	2406	3	US-08-463-074B-1 Sequence 1, Appli
	299	17	0.3	2406	3	US-08-465-585C-1 Sequence 1, Appli
	300	17	0.3	2406	3	US-08-652-446-1 Sequence 1, Appli
	301	17	0.3	2481	2	US-08-958-642-3 Sequence 3, Appli
	302	17	0.3	2481	3	US-08-778-394-1 Sequence 1, Appli
	303	17	0.3	2481	3	US-08-778-423A-3 Sequence 3, Appli
c	304	17	0.3	2511	1	US-08-363-560-1 Sequence 1, Appli
	305	17	0.3	2584	3	US-08-758-662-8 Sequence 8, Appli
	306	17	0.3	2784	1	US-08-471-454-1 Sequence 1, Appli
	307	17	0.3	2784	2	US-08-466-974-1 Sequence 1, Appli
	308	17	0.3	2784	2	US-08-471-453-1 Sequence 1, Appli
c	309	17	0.3	2789	1	US-07-998-003A-104 Sequence 104, App
c	310	17	0.3	2789	1	US-08-453-274B-104 Sequence 104, App
c	311	17	0.3	2789	1	US-08-453-695A-104 Sequence 104, App
c	312	17	0.3	2789	1	US-08-268-161A-104 Sequence 104, App

c	313	17	0.3	2789	2	US-08-453-702A-104 Sequence 104, App
c	314	17	0.3	2789	4	PCT-US93-12588-104 Sequence 104, App
c	315	17	0.3	2789	4	PCT-US95-08071-104 Sequence 104, App
	316	17	0.3	2840	4	PCT-US93-06251-64 Sequence 64, Appl
	317	17	0.3	2940	2	US-08-418-848A-3 Sequence 3, Appli
	318	17	0.3	3032	3	US-08-990-140-1 Sequence 1, Appli
	319	17	0.3	3043	1	US-08-188-228-43 Sequence 43, Appl
	320	17	0.3	3043	1	US-08-332-638-43 Sequence 43, Appl
	321	17	0.3	3138	1	US-07-867-106-4 Sequence 4, Appli
	322	17	0.3	3195	2	US-08-951-648-5 Sequence 5, Appli
	323	17	0.3	3195	3	US-09-174-437-5 Sequence 5, Appli
c	324	17	0.3	3296	3	US-08-335-844A-14 Sequence 14, Appl
	325	17	0.3	3312	1	US-08-049-473-1 Sequence 1, Appli
	326	17	0.3	3312	1	US-08-312-648-1 Sequence 1, Appli
	327	17	0.3	3312	4	PCT-US94-04190-1 Sequence 1, Appli
	328	17	0.3	3336	1	US-08-368-803-6 Sequence 6, Appli
c	329	17	0.3	3395	5	5223424-11 Patent No. 5223424
	330	17	0.3	3785	1	US-08-485-618-98 Sequence 98, Appl
	331	17	0.3	3785	1	US-08-605-672-98 Sequence 98, Appl
	332	17	0.3	3785	2	US-08-482-293A-98 Sequence 98, Appl
	333	17	0.3	3785	2	US-08-943-363-98 Sequence 98, Appl
	334	17	0.3	3956	1	US-08-485-618-97 Sequence 97, Appl
	335	17	0.3	3956	1	US-08-605-672-97 Sequence 97, Appl
	336	17	0.3	3956	2	US-08-482-293A-97 Sequence 97, Appl
	337	17	0.3	3956	2	US-08-943-363-97 Sequence 97, Appl
c	338	17	0.3	3958	1	US-08-435-933-5 Sequence 5, Appli
c	339	17	0.3	3958	4	PCT-US96-06035-5 Sequence 5, Appli
	340	17	0.3	4322	1	US-08-673-789-1 Sequence 1, Appli
c	341	17	0.3	4467	1	US-08-565-907A-1 Sequence 1, Appli
c	342	17	0.3	4467	2	US-08-910-551B-1 Sequence 1, Appli
c	343	17	0.3	4467	2	US-08-909-425A-1 Sequence 1, Appli
c	344	17	0.3	4471	2	US-08-615-942A-1 Sequence 1, Appli
c	345	17	0.3	4599	1	US-08-431-080-27 Sequence 27, Appl
c	346	17	0.3	4599	2	US-08-938-534-27 Sequence 27, Appl
c	347	17	0.3	4705	1	US-07-998-003A-96 Sequence 96, Appl
c	348	17	0.3	4705	1	US-08-453-274B-96 Sequence 96, Appl
c	349	17	0.3	4705	1	US-08-453-695A-96 Sequence 96, Appl
c	350	17	0.3	4705	1	US-08-268-161A-96 Sequence 96, Appl
c	351	17	0.3	4705	2	US-08-453-702A-96 Sequence 96, Appl
c	352	17	0.3	4705	4	PCT-US93-12588-96 Sequence 96, Appl
c	353	17	0.3	4705	4	PCT-US95-08071-96 Sequence 96, Appl
c	354	17	0.3	4977	1	US-08-030-096-7 Sequence 7, Appli
	355	17	0.3	5340	3	US-08-492-459-9 Sequence 9, Appli
	356	17	0.3	5340	3	US-08-423-752-9 Sequence 9, Appli
	357	17	0.3	5340	3	US-08-945-994-2 Sequence 2, Appli
	358	17	0.3	5426	2	US-08-663-566A-3 Sequence 3, Appli
	359	17	0.3	5426	2	US-08-023-610-3 Sequence 3, Appli
	360	17	0.3	5426	2	US-08-288-065A-3 Sequence 3, Appli
	361	17	0.3	5426	2	US-08-362-240A-3 Sequence 3, Appli
	362	17	0.3	5426	4	PCT-US95-10245-3 Sequence 3, Appli
	363	17	0.3	6045	3	US-08-675-566-18 Sequence 18, Appl

c 364	17	0.3	6202	1	US-08-484-101B-41	Sequence 41, Appl
365	17	0.3	6244	3	US-08-675-566-17	Sequence 17, Appl
366	17	0.3	6447	3	US-08-675-566-16	Sequence 16, Appl
367	17	0.3	6578	3	US-08-675-566-4	Sequence 4, Appli
368	17	0.3	6612	3	US-08-675-566-15	Sequence 15, Appl
c 369	17	0.3	6763	2	US-08-756-506-23	Sequence 23, Appl
370	17	0.3	6958	3	US-08-675-566-2	Sequence 2, Appli
371	17	0.3	6994	3	US-08-675-566-1	Sequence 1, Appli
372	17	0.3	7001	3	US-08-675-566-3	Sequence 3, Appli
373	17	0.3	7102	3	US-09-138-024-20	Sequence 20, Appl
c 374	17	0.3	7225	2	US-08-286-819A-15	Sequence 15, Appl
c 375	17	0.3	7225	3	US-08-980-357-15	Sequence 15, Appl
376	17	0.3	7333	3	US-09-138-024-21	Sequence 21, Appl
377	17	0.3	7859	1	US-07-854-596B-4	Sequence 4, Appli
c 378	17	0.3	7859	1	US-07-854-596B-4	Sequence 4, Appli
379	17	0.3	7859	2	US-08-450-905B-15	Sequence 15, Appl
c 380	17	0.3	7859	2	US-08-450-905B-15	Sequence 15, Appl
381	17	0.3	7859	3	US-07-982-759F-15	Sequence 15, Appl
c 382	17	0.3	7859	3	US-07-982-759F-15	Sequence 15, Appl
c 383	17	0.3	10807	1	US-08-206-176-7	Sequence 7, Appli
c 384	17	0.3	10807	2	US-08-756-506-5	Sequence 5, Appli
c 385	17	0.3	10851	2	US-08-286-819A-16	Sequence 16, Appl
c 386	17	0.3	10851	3	US-08-980-357-16	Sequence 16, Appl
387	17	0.3	10968	2	US-08-680-327-2	Sequence 2, Appli
c 388	17	0.3	12124	1	US-08-181-271A-36	Sequence 36, Appl
c 389	17	0.3	12124	1	US-08-449-315-36	Sequence 36, Appl
c 390	17	0.3	12124	1	US-08-444-803-36	Sequence 36, Appl
c 391	17	0.3	12124	1	US-08-449-043-36	Sequence 36, Appl
c 392	17	0.3	12124	1	US-08-456-265A-36	Sequence 36, Appl
c 393	17	0.3	12124	1	US-08-455-416-36	Sequence 36, Appl
c 394	17	0.3	12124	1	US-08-455-244-36	Sequence 36, Appl
c 395	17	0.3	12124	1	US-08-454-876-36	Sequence 36, Appl
c 396	17	0.3	12124	2	US-08-457-364-36	Sequence 36, Appl
c 397	17	0.3	12124	2	US-08-456-262-36	Sequence 36, Appl
c 398	17	0.3	12124	2	US-08-456-240-36	Sequence 36, Appl
c 399	17	0.3	12124	2	US-08-455-736-36	Sequence 36, Appl
c 400	17	0.3	12124	2	US-08-971-217-36	Sequence 36, Appl
401	17	0.3	17041	1	US-08-076-011-1	Sequence 1, Appli
402	17	0.3	43676	3	US-09-356-952-12	Sequence 12, Appl
c 403	17	0.3	68750	3	US-09-335-409-1	Sequence 1, Appli
c 404	17	0.3	87350	3	US-08-781-891-79	Sequence 79, Appl
c 405	17	0.3	152331	3	US-09-128-155-16	Sequence 16, Appl
406	17	0.3	176373	3	US-09-128-155-17	Sequence 17, Appl
c 407	16	0.3	30	3	US-08-771-781-2	Sequence 2, Appli
c 408	16	0.3	36	1	US-08-616-732A-19	Sequence 19, Appl
c 409	16	0.3	36	3	US-09-037-742B-19	Sequence 19, Appl
c 410	16	0.3	40	3	US-08-688-514-14	Sequence 14, Appl
411	16	0.3	50	5	5457025-3	Patent No. 5457025
412	16	0.3	54	2	US-08-771-624B-4	Sequence 4, Appli
413	16	0.3	55	2	US-08-771-624B-8	Sequence 8, Appli
414	16	0.3	68	2	US-08-600-783-18	Sequence 18, Appl

415	16	0.3	97	1	US-08-120-827-87	Sequence 87, Appl
416	16	0.3	97	1	US-08-478-675-87	Sequence 87, Appl
417	16	0.3	109	1	US-08-120-827-81	Sequence 81, Appl
418	16	0.3	109	1	US-08-478-675-81	Sequence 81, Appl
419	16	0.3	115	1	US-08-120-827-88	Sequence 88, Appl
420	16	0.3	115	1	US-08-478-675-88	Sequence 88, Appl
421	16	0.3	197	2	US-08-520-670A-31	Sequence 31, Appl
422	16	0.3	199	1	US-08-330-108-4	Sequence 4, Appli
423	16	0.3	199	4	PCT-US92-10087-4	Sequence 4, Appli
424	16	0.3	246	3	US-08-765-332-116	Sequence 116, App
425	16	0.3	258	2	US-08-600-783-17	Sequence 17, Appl
426	16	0.3	275	1	US-08-215-084A-2	Sequence 2, Appli
427	16	0.3	275	1	US-08-463-212-2	Sequence 2, Appli
428	16	0.3	275	1	US-08-463-211-2	Sequence 2, Appli
429	16	0.3	285	2	US-08-630-822A-85	Sequence 85, Appl
430	16	0.3	285	2	US-09-005-069-85	Sequence 85, Appl
431	16	0.3	306	1	US-08-627-706-8	Sequence 8, Appli
432	16	0.3	309	3	US-08-581-148C-2	Sequence 2, Appli
433	16	0.3	341	3	US-09-029-267-16	Sequence 16, Appl
434	16	0.3	349	1	US-08-220-606B-55	Sequence 55, Appl
435	16	0.3	363	3	US-08-765-332-119	Sequence 119, App
c 436	16	0.3	383	3	US-08-978-454-3	Sequence 3, Appli
c 437	16	0.3	434	1	US-08-480-784-34	Sequence 34, Appl
c 438	16	0.3	434	1	US-08-483-553-34	Sequence 34, Appl
c 439	16	0.3	434	1	US-08-487-002-34	Sequence 34, Appl
c 440	16	0.3	434	1	US-08-483-554B-34	Sequence 34, Appl
c 441	16	0.3	434	1	US-08-488-011B-34	Sequence 34, Appl
c 442	16	0.3	434	3	US-08-850-727-34	Sequence 34, Appl
c 443	16	0.3	434	4	PCT-US95-10202-34	Sequence 34, Appl
c 444	16	0.3	434	4	PCT-US95-10203-34	Sequence 34, Appl
c 445	16	0.3	434	4	PCT-US95-10220-34	Sequence 34, Appl
446	16	0.3	452	1	US-07-662-198B-1	Sequence 1, Appli
447	16	0.3	452	1	US-08-322-742-1	Sequence 1, Appli
448	16	0.3	468	3	US-08-978-454-1	Sequence 1, Appli
449	16	0.3	498	1	US-07-885-970A-8	Sequence 8, Appli
450	16	0.3	498	1	US-08-298-687A-8	Sequence 8, Appli
c 451	16	0.3	498	1	US-08-530-797-7	Sequence 7, Appli
452	16	0.3	498	1	US-08-298-829-8	Sequence 8, Appli
453	16	0.3	498	2	US-08-791-849A-1	Sequence 1, Appli
c 454	16	0.3	498	2	US-08-787-335-7	Sequence 7, Appli
455	16	0.3	500	1	US-08-627-706-9	Sequence 9, Appli
456	16	0.3	516	2	US-08-923-738-3	Sequence 3, Appli
c 457	16	0.3	518	1	US-08-485-284A-2	Sequence 2, Appli
458	16	0.3	519	2	US-08-923-738-1	Sequence 1, Appli
c 459	16	0.3	529	3	US-08-545-809A-44	Sequence 44, Appl
c 460	16	0.3	530	2	US-08-346-429-5	Sequence 5, Appli
461	16	0.3	536	1	US-08-329-704-1	Sequence 1, Appli
462	16	0.3	536	2	US-08-472-604-1	Sequence 1, Appli
463	16	0.3	536	2	US-08-486-117-1	Sequence 1, Appli
464	16	0.3	548	3	US-09-036-355A-1	Sequence 1, Appli
465	16	0.3	570	2	US-08-633-682-1	Sequence 1, Appli

466	16	0.3	570	3	US-08-936-772-1	Sequence 1, Appli
467	16	0.3	585	3	US-08-581-148C-11	Sequence 11, Appl
468	16	0.3	594	1	US-08-289-458-1	Sequence 1, Appli
469	16	0.3	594	2	US-08-761-549-1	Sequence 1, Appli
470	16	0.3	619	1	US-07-847-010-9	Sequence 9, Appli
471	16	0.3	623	5	5240847-1	Patent No. 5240847
472	16	0.3	633	3	US-09-050-603A-36	Sequence 36, Appl
473	16	0.3	633	3	US-09-102-420B-36	Sequence 36, Appl
474	16	0.3	641	2	US-08-522-421-6	Sequence 6, Appli
c 475	16	0.3	648	2	US-08-935-396-9	Sequence 9, Appli
c 476	16	0.3	663	2	US-08-943-087-59	Sequence 59, Appl
477	16	0.3	686	1	US-08-591-498-15	Sequence 15, Appl
478	16	0.3	700	1	US-07-846-992-1	Sequence 1, Appli
479	16	0.3	700	1	US-08-469-555-1	Sequence 1, Appli
480	16	0.3	704	3	US-09-048-889-7	Sequence 7, Appli
481	16	0.3	707	2	US-08-850-910A-40	Sequence 40, Appl
482	16	0.3	720	3	US-08-946-026-14	Sequence 14, Appl
483	16	0.3	747	3	US-09-393-877-8	Sequence 8, Appli
c 484	16	0.3	750	2	US-08-602-359A-29	Sequence 29, Appl
485	16	0.3	751	1	US-08-592-936B-10	Sequence 10, Appl
486	16	0.3	751	1	US-08-788-928A-16	Sequence 16, Appl
487	16	0.3	751	2	US-09-111-573-10	Sequence 10, Appl
488	16	0.3	753	3	US-09-393-877-9	Sequence 9, Appli
489	16	0.3	756	2	US-08-530-165-1	Sequence 1, Appli
490	16	0.3	765	1	US-08-173-510B-102	Sequence 102, App
491	16	0.3	765	1	US-08-458-218-100	Sequence 100, App
492	16	0.3	765	2	US-08-450-497-102	Sequence 102, App
493	16	0.3	767	1	US-08-184-604-1	Sequence 1, Appli
494	16	0.3	794	3	US-09-188-930-41	Sequence 41, Appl
495	16	0.3	797	3	US-09-188-930-35	Sequence 35, Appl
496	16	0.3	804	1	US-08-411-635-1	Sequence 1, Appli
497	16	0.3	804	4	PCT-US94-10644-1	Sequence 1, Appli
c 498	16	0.3	815	2	US-08-978-458-3	Sequence 3, Appli
499	16	0.3	842	1	US-08-208-181A-23	Sequence 23, Appl
500	16	0.3	846	3	US-09-154-874-1	Sequence 1, Appli
501	16	0.3	853	3	US-09-081-180-1	Sequence 1, Appli
502	16	0.3	853	3	US-09-040-786-1	Sequence 1, Appli
503	16	0.3	855	5	5185441-40	Patent No. 5185441
504	16	0.3	855	5	5223394-3	Patent No. 5223394
505	16	0.3	863	1	US-07-940-861-11	Sequence 11, Appl
506	16	0.3	863	1	US-08-459-512-11	Sequence 11, Appl
507	16	0.3	863	2	US-08-459-657-11	Sequence 11, Appl
508	16	0.3	863	2	US-08-460-132-11	Sequence 11, Appl
509	16	0.3	863	4	PCT-US92-02050-11	Sequence 11, Appl
510	16	0.3	863	5	5185441-35	Patent No. 5185441
511	16	0.3	863	5	5223394-5	Patent No. 5223394
512	16	0.3	866	5	5185431-3	Patent No. 5185431
513	16	0.3	876	2	US-08-937-972-2	Sequence 2, Appli
514	16	0.3	880	1	US-08-616-368A-7	Sequence 7, Appli
515	16	0.3	880	3	US-09-054-298-7	Sequence 7, Appli
516	16	0.3	896	3	US-08-581-148C-8	Sequence 8, Appli

517	16	0.3	897	2	US-08-630-822A-59	Sequence 59, Appl
518	16	0.3	897	2	US-09-005-069-59	Sequence 59, Appl
519	16	0.3	912	2	US-08-993-228-7	Sequence 7, Appli
520	16	0.3	913	3	US-08-651-136C-13	Sequence 13, Appl
521	16	0.3	917	2	US-08-951-822-1	Sequence 1, Appli
522	16	0.3	925	1	US-08-544-900-3	Sequence 3, Appli
523	16	0.3	925	4	PCT-US95-07874-2	Sequence 2, Appli
524	16	0.3	927	2	US-07-690-192-1	Sequence 1, Appli
525	16	0.3	943	2	US-08-303-569B-4	Sequence 4, Appli
526	16	0.3	943	2	US-08-116-247-4	Sequence 4, Appli
527	16	0.3	953	1	US-08-197-793-1	Sequence 1, Appli
528	16	0.3	953	2	US-08-636-176-1	Sequence 1, Appli
529	16	0.3	953	4	PCT-US95-01618-1	Sequence 1, Appli
530	16	0.3	969	2	US-08-366-486A-27	Sequence 27, Appl
c 531	16	0.3	974	3	US-08-675-885-3	Sequence 3, Appli
532	16	0.3	1007	3	US-09-248-335-31	Sequence 31, Appl
533	16	0.3	1009	5	5223394-8	Patent No. 5223394
534	16	0.3	1023	1	US-08-698-551-7	Sequence 7, Appli
535	16	0.3	1023	2	US-08-602-228-7	Sequence 7, Appli
536	16	0.3	1023	2	US-08-649-341A-7	Sequence 7, Appli
537	16	0.3	1023	2	US-08-494-440B-7	Sequence 7, Appli
538	16	0.3	1023	2	US-08-533-901B-7	Sequence 7, Appli
539	16	0.3	1023	2	US-08-839-032A-7	Sequence 7, Appli
540	16	0.3	1023	2	US-08-839-031A-7	Sequence 7, Appli
541	16	0.3	1023	3	US-08-554-385-4	Sequence 4, Appli
542	16	0.3	1023	4	PCT-US92-01196-5	Sequence 5, Appli
543	16	0.3	1023	4	PCT-US95-12724-7	Sequence 7, Appli
c 544	16	0.3	1038	1	US-08-680-726A-53	Sequence 53, Appl
c 545	16	0.3	1038	3	US-09-092-409-53	Sequence 53, Appl
546	16	0.3	1045	2	US-09-014-969-6	Sequence 6, Appli
547	16	0.3	1051	2	US-08-865-273-1	Sequence 1, Appli
548	16	0.3	1054	1	US-08-152-485-1	Sequence 1, Appli
549	16	0.3	1054	1	US-08-463-089-1	Sequence 1, Appli
550	16	0.3	1054	1	US-08-461-360A-1	Sequence 1, Appli
551	16	0.3	1054	1	US-08-461-359-1	Sequence 1, Appli
552	16	0.3	1054	4	PCT-US94-12904-1	Sequence 1, Appli
553	16	0.3	1056	3	US-08-755-587-28	Sequence 28, Appl
554	16	0.3	1067	2	US-08-507-598-1	Sequence 1, Appli
555	16	0.3	1067	2	US-08-507-750-1	Sequence 1, Appli
556	16	0.3	1067	3	US-08-764-522A-1	Sequence 1, Appli
557	16	0.3	1067	3	US-08-764-528-1	Sequence 1, Appli
558	16	0.3	1067	3	US-08-872-859-1	Sequence 1, Appli
559	16	0.3	1068	3	US-09-248-335-73	Sequence 73, Appl
560	16	0.3	1073	3	US-08-911-423-1	Sequence 1, Appli
561	16	0.3	1078	5	5223394-10	Patent No. 5223394
562	16	0.3	1096	3	US-09-136-073-1	Sequence 1, Appli
563	16	0.3	1101	1	US-08-525-505A-1	Sequence 1, Appli
564	16	0.3	1105	3	US-08-961-083-75	Sequence 75, Appl
565	16	0.3	1107	1	US-08-181-271A-8	Sequence 8, Appli
566	16	0.3	1107	1	US-08-449-315-8	Sequence 8, Appli
567	16	0.3	1107	1	US-08-444-803-8	Sequence 8, Appli

	568	16	0.3	1107	1	US-08-449-043-8 Sequence 8, Appli
	569	16	0.3	1107	1	US-08-456-265A-8 Sequence 8, Appli
	570	16	0.3	1107	1	US-08-455-416-8 Sequence 8, Appli
	571	16	0.3	1107	1	US-08-455-244-8 Sequence 8, Appli
	572	16	0.3	1107	1	US-08-454-876-8 Sequence 8, Appli
	573	16	0.3	1107	2	US-08-457-364-8 Sequence 8, Appli
	574	16	0.3	1107	2	US-08-456-262-8 Sequence 8, Appli
	575	16	0.3	1107	2	US-08-456-240-8 Sequence 8, Appli
	576	16	0.3	1107	2	US-08-455-736-8 Sequence 8, Appli
	577	16	0.3	1107	2	US-08-971-217-8 Sequence 8, Appli
	578	16	0.3	1108	1	US-08-181-271A-13 Sequence 13, Appl
	579	16	0.3	1108	1	US-08-449-315-13 Sequence 13, Appl
	580	16	0.3	1108	1	US-08-444-803-13 Sequence 13, Appl
	581	16	0.3	1108	1	US-08-449-043-13 Sequence 13, Appl
	582	16	0.3	1108	1	US-08-456-265A-13 Sequence 13, Appl
	583	16	0.3	1108	1	US-08-455-416-13 Sequence 13, Appl
	584	16	0.3	1108	1	US-08-455-244-13 Sequence 13, Appl
	585	16	0.3	1108	1	US-08-454-876-13 Sequence 13, Appl
	586	16	0.3	1108	2	US-08-457-364-13 Sequence 13, Appl
	587	16	0.3	1108	2	US-08-456-262-13 Sequence 13, Appl
	588	16	0.3	1108	2	US-08-456-240-13 Sequence 13, Appl
	589	16	0.3	1108	2	US-08-455-736-13 Sequence 13, Appl
	590	16	0.3	1108	2	US-08-971-217-13 Sequence 13, Appl
	591	16	0.3	1113	2	US-08-907-492A-3 Sequence 3, Appli
	592	16	0.3	1115	1	US-08-190-802A-19 Sequence 19, Appl
	593	16	0.3	1120	3	US-08-884-324-2 Sequence 2, Appli
c	594	16	0.3	1129	4	PCT-US93-01676A-3 Sequence 3, Appli
c	595	16	0.3	1132	2	US-08-928-692-50 Sequence 50, Appl
	596	16	0.3	1133	3	US-08-983-409-5 Sequence 5, Appli
	597	16	0.3	1140	2	US-08-698-805-5 Sequence 5, Appli
	598	16	0.3	1144	3	US-09-124-758-3 Sequence 3, Appli
	599	16	0.3	1152	1	US-08-315-831A-13 Sequence 13, Appl
	600	16	0.3	1152	1	US-08-662-318-13 Sequence 13, Appl
	601	16	0.3	1152	4	PCT-US95-12509-13 Sequence 13, Appl
	602	16	0.3	1195	1	US-08-181-271A-14 Sequence 14, Appl
	603	16	0.3	1195	1	US-08-449-315-14 Sequence 14, Appl
	604	16	0.3	1195	1	US-08-444-803-14 Sequence 14, Appl
	605	16	0.3	1195	1	US-08-449-043-14 Sequence 14, Appl
	606	16	0.3	1195	1	US-08-456-265A-14 Sequence 14, Appl
	607	16	0.3	1195	1	US-08-455-416-14 Sequence 14, Appl
	608	16	0.3	1195	1	US-08-455-244-14 Sequence 14, Appl
	609	16	0.3	1195	1	US-08-454-876-14 Sequence 14, Appl
	610	16	0.3	1195	2	US-08-457-364-14 Sequence 14, Appl
	611	16	0.3	1195	2	US-08-456-262-14 Sequence 14, Appl
	612	16	0.3	1195	2	US-08-456-240-14 Sequence 14, Appl
	613	16	0.3	1195	2	US-08-455-736-14 Sequence 14, Appl
	614	16	0.3	1195	2	US-08-971-217-14 Sequence 14, Appl
	615	16	0.3	1217	3	US-09-225-244-1 Sequence 1, Appli
	616	16	0.3	1218	3	US-09-154-874-7 Sequence 7, Appli
	617	16	0.3	1244	2	US-08-204-288-3 Sequence 3, Appli
	618	16	0.3	1255	1	US-08-518-878B-38 Sequence 38, Appl

	619	16	0.3	1255	1	US-08-294-522B-38 Sequence 38, Appl
	620	16	0.3	1255	2	US-08-470-868A-38 Sequence 38, Appl
	621	16	0.3	1257	1	US-08-487-748A-11 Sequence 11, Appl
	622	16	0.3	1257	3	US-08-480-070C-11 Sequence 11, Appl
	623	16	0.3	1257	3	US-08-829-525-11 Sequence 11, Appl
	624	16	0.3	1258	1	US-08-335-518-1 Sequence 1, Appli
	625	16	0.3	1258	3	US-08-988-054-1 Sequence 1, Appli
	626	16	0.3	1267	1	US-08-140-797-1 Sequence 1, Appli
	627	16	0.3	1267	1	US-08-486-670A-1 Sequence 1, Appli
	628	16	0.3	1279	3	US-09-248-335-25 Sequence 25, Appl
	629	16	0.3	1293	1	US-08-317-522A-6 Sequence 6, Appli
	630	16	0.3	1293	1	US-08-439-818A-6 Sequence 6, Appli
	631	16	0.3	1293	1	US-08-751-965-6 Sequence 6, Appli
	632	16	0.3	1293	2	US-08-738-975-6 Sequence 6, Appli
	633	16	0.3	1293	2	US-08-728-626-6 Sequence 6, Appli
	634	16	0.3	1293	3	US-08-808-599A-6 Sequence 6, Appli
	635	16	0.3	1299	1	US-08-723-938-4 Sequence 4, Appli
	636	16	0.3	1299	2	US-09-080-538-4 Sequence 4, Appli
	637	16	0.3	1299	3	US-08-793-035-3 Sequence 3, Appli
	638	16	0.3	1315	2	US-08-578-592-4 Sequence 4, Appli
	639	16	0.3	1315	3	US-09-185-111-4 Sequence 4, Appli
	640	16	0.3	1316	1	US-07-755-461A-12 Sequence 12, Appl
	641	16	0.3	1316	1	US-08-315-831A-12 Sequence 12, Appl
	642	16	0.3	1316	1	US-08-662-318-12 Sequence 12, Appl
	643	16	0.3	1316	4	PCT-US95-12509-12 Sequence 12, Appl
	644	16	0.3	1317	1	US-08-151-848-45 Sequence 45, Appl
	645	16	0.3	1317	3	US-09-299-843A-45 Sequence 45, Appl
	646	16	0.3	1317	4	PCT-US93-11153-45 Sequence 45, Appl
c	647	16	0.3	1320	1	US-08-220-151-18 Sequence 18, Appl
c	648	16	0.3	1320	1	US-08-413-118-18 Sequence 18, Appl
	649	16	0.3	1320	1	US-08-419-414-1 Sequence 1, Appli
c	650	16	0.3	1320	3	US-08-473-446-18 Sequence 18, Appl
	651	16	0.3	1339	1	US-07-936-163-2 Sequence 2, Appli
	652	16	0.3	1339	3	US-08-468-056B-3 Sequence 3, Appli
	653	16	0.3	1339	3	US-08-468-059A-3 Sequence 3, Appli
	654	16	0.3	1351	2	US-08-401-068-13 Sequence 13, Appl
	655	16	0.3	1351	2	US-08-846-338-13 Sequence 13, Appl
	656	16	0.3	1357	5	5340934-7 Patent No. 5340934
	657	16	0.3	1359	3	US-09-188-930-34 Sequence 34, Appl
	658	16	0.3	1373	3	US-08-725-758A-1 Sequence 1, Appli
c	659	16	0.3	1387	3	US-08-868-594-1 Sequence 1, Appli
	660	16	0.3	1393	1	US-08-174-467-18 Sequence 18, Appl
	661	16	0.3	1393	3	US-08-452-071-18 Sequence 18, Appl
c	662	16	0.3	1400	1	US-08-464-164-1 Sequence 1, Appli
c	663	16	0.3	1400	1	US-08-338-057-1 Sequence 1, Appli
c	664	16	0.3	1400	2	US-08-668-416-1 Sequence 1, Appli
	665	16	0.3	1407	3	US-08-581-148C-19 Sequence 19, Appl
	666	16	0.3	1408	1	US-08-447-554-3 Sequence 3, Appli
	667	16	0.3	1408	1	US-08-448-160-3 Sequence 3, Appli
	668	16	0.3	1408	1	US-08-254-404-21 Sequence 21, Appl
	669	16	0.3	1408	2	US-08-327-451E-21 Sequence 21, Appl

	670	16	0.3	1408	2	US-08-458-109-21 Sequence 21, Appl
c	671	16	0.3	1409	2	US-08-598-873-69 Sequence 69, Appl
c	672	16	0.3	1415	1	US-08-413-118-126 Sequence 126, App
c	673	16	0.3	1415	3	US-08-473-446-126 Sequence 126, App
c	674	16	0.3	1433	2	US-08-428-243-6 Sequence 6, Appli
c	675	16	0.3	1433	4	PCT-US93-10301-6 Sequence 6, Appli
	676	16	0.3	1454	1	US-08-467-155A-2 Sequence 2, Appli
	677	16	0.3	1454	2	US-08-628-198-2 Sequence 2, Appli
	678	16	0.3	1454	3	US-09-201-038-2 Sequence 2, Appli
	679	16	0.3	1454	4	PCT-US96-07343-2 Sequence 2, Appli
	680	16	0.3	1457	2	US-08-484-993B-23 Sequence 23, Appl
	681	16	0.3	1457	2	US-08-484-158B-23 Sequence 23, Appl
	682	16	0.3	1457	2	US-08-484-596A-23 Sequence 23, Appl
	683	16	0.3	1457	2	US-08-480-150A-23 Sequence 23, Appl
	684	16	0.3	1457	3	US-08-458-731-23 Sequence 23, Appl
	685	16	0.3	1457	3	US-08-149-223A-23 Sequence 23, Appl
	686	16	0.3	1460	5	5225537-3 Patent No. 5225537
	687	16	0.3	1479	3	US-08-946-914-5 Sequence 5, Appli
c	688	16	0.3	1480	2	US-08-454-557C-38 Sequence 38, Appl
c	689	16	0.3	1480	2	US-08-340-426D-38 Sequence 38, Appl
c	690	16	0.3	1480	2	US-08-450-673C-38 Sequence 38, Appl
c	691	16	0.3	1480	4	PCT-US95-17111A-38 Sequence 38, Appl
	692	16	0.3	1492	3	US-08-350-468-7 Sequence 7, Appli
	693	16	0.3	1493	2	US-08-820-170A-9 Sequence 9, Appli
	694	16	0.3	1493	3	US-09-055-699-9 Sequence 9, Appli
	695	16	0.3	1493	3	US-09-273-565-9 Sequence 9, Appli
	696	16	0.3	1502	3	US-09-046-736-3 Sequence 3, Appli
	697	16	0.3	1504	2	US-08-850-910A-17 Sequence 17, Appl
c	698	16	0.3	1505	2	US-08-428-243-5 Sequence 5, Appli
c	699	16	0.3	1505	4	PCT-US93-10301-5 Sequence 5, Appli
	700	16	0.3	1507	2	US-08-850-910A-38 Sequence 38, Appl
	701	16	0.3	1537	5	5504194-1 Patent No. 5504194
	702	16	0.3	1558	1	US-08-487-823B-1 Sequence 1, Appli
	703	16	0.3	1558	2	US-08-997-040-1 Sequence 1, Appli
	704	16	0.3	1558	2	US-09-203-237-1 Sequence 1, Appli
	705	16	0.3	1559	2	US-08-417-174-1 Sequence 1, Appli
	706	16	0.3	1559	2	US-08-231-565A-1 Sequence 1, Appli
	707	16	0.3	1559	2	US-09-007-961-1 Sequence 1, Appli
	708	16	0.3	1584	4	PCT-US93-04392-4 Sequence 4, Appli
	709	16	0.3	1584	4	PCT-US93-04392-7 Sequence 7, Appli
	710	16	0.3	1584	4	PCT-US93-04392-10 Sequence 10, Appl
	711	16	0.3	1584	4	PCT-US93-04392-13 Sequence 13, Appl
	712	16	0.3	1586	1	US-07-841-646-18 Sequence 18, Appl
	713	16	0.3	1586	1	US-08-147-023-18 Sequence 18, Appl
	714	16	0.3	1586	1	US-08-447-570-18 Sequence 18, Appl
	715	16	0.3	1586	2	US-08-449-700-18 Sequence 18, Appl
	716	16	0.3	1586	2	US-08-449-699A-18 Sequence 18, Appl
	717	16	0.3	1596	2	US-08-807-861A-38 Sequence 38, Appl
	718	16	0.3	1596	3	US-09-210-681-38 Sequence 38, Appl
	719	16	0.3	1596	3	US-08-946-719A-38 Sequence 38, Appl
	720	16	0.3	1608	2	US-08-424-224-1 Sequence 1, Appli

	721	16	0.3	1608	4	PCT-US94-02891-68 Sequence 68, Appl
	722	16	0.3	1630	3	US-08-468-856B-2 Sequence 2, Appli
	723	16	0.3	1630	3	US-08-468-859A-2 Sequence 2, Appli
	724	16	0.3	1642	3	US-08-993-359-27 Sequence 27, Appl
c	725	16	0.3	1647	1	US-08-198-446B-9 Sequence 9, Appli
c	726	16	0.3	1647	2	US-08-870-693-9 Sequence 9, Appli
	727	16	0.3	1654	3	US-08-991-426-3 Sequence 3, Appli
	728	16	0.3	1654	3	US-09-143-470-3 Sequence 3, Appli
	729	16	0.3	1661	1	US-08-318-492-3 Sequence 3, Appli
	730	16	0.3	1661	1	US-08-707-340-3 Sequence 3, Appli
	731	16	0.3	1661	2	US-08-994-578-3 Sequence 3, Appli
	732	16	0.3	1679	1	US-08-202,056-6 Sequence 6, Appli
	733	16	0.3	1679	1	US-08-076-093A-5 Sequence 5, Appli
	734	16	0.3	1679	1	US-08-701-265-5 Sequence 5, Appli
	735	16	0.3	1679	2	US-08-284-586-5 Sequence 5, Appli
	736	16	0.3	1679	2	US-08-805-478-5 Sequence 5, Appli
	737	16	0.3	1679	2	US-08-802-627A-5 Sequence 5, Appli
	738	16	0.3	1679	2	US-08-801-238-5 Sequence 5, Appli
	739	16	0.3	1679	2	US-08-801-228-5 Sequence 5, Appli
	740	16	0.3	1679	3	US-09-104-296-5 Sequence 5, Appli
	741	16	0.3	1679	4	PCT-US94-06380-3 Sequence 3, Appli
	742	16	0.3	1688	2	US-08-439-814-2 Sequence 2, Appli
	743	16	0.3	1691	1	US-08-472-028A-5 Sequence 5, Appli
	744	16	0.3	1691	2	US-08-808-931-5 Sequence 5, Appli
	745	16	0.3	1691	3	US-08-808-323-5 Sequence 5, Appli
	746	16	0.3	1691	3	US-09-050-603A-5 Sequence 5, Appli
	747	16	0.3	1691	3	US-09-102-420B-5 Sequence 5, Appli
	748	16	0.3	1703	1	US-07-868-353A-11 Sequence 11, Appl
	749	16	0.3	1703	1	US-08-407-804-20 Sequence 20, Appl
	750	16	0.3	1703	3	US-09-124-807-20 Sequence 20, Appl
	751	16	0.3	1712	2	US-08-632-598-1 Sequence 1, Appli
	752	16	0.3	1716	2	US-08-954-333-9 Sequence 9, Appli
	753	16	0.3	1723	1	US-08-417-492-1 Sequence 1, Appli
	754	16	0.3	1743	3	US-09-032-365A-18 Sequence 18, Appl
	755	16	0.3	1745	1	US-08-464-523B-1 Sequence 1, Appli
	756	16	0.3	1747	1	US-08-202-389-1 Sequence 1, Appli
	757	16	0.3	1755	1	US-07-912-900-29 Sequence 29, Appl
	758	16	0.3	1755	1	US-08-285-309-29 Sequence 29, Appl
	759	16	0.3	1755	2	US-08-502-046-29 Sequence 29, Appl
	760	16	0.3	1758	2	US-08-630-822A-69 Sequence 69, Appl
	761	16	0.3	1758	2	US-09-005-069-69 Sequence 69, Appl
	762	16	0.3	1772	2	US-08-960-022-13 Sequence 13, Appl
	763	16	0.3	1777	2	US-08-937-466-5 Sequence 5, Appli
	764	16	0.3	1777	2	US-09-172-528-5 Sequence 5, Appli
	765	16	0.3	1777	3	US-09-318-199-5 Sequence 5, Appli
c	766	16	0.3	1780	2	US-08-480-229C-11 Sequence 11, Appl
c	767	16	0.3	1780	2	US-08-659-235C-11 Sequence 11, Appl
	768	16	0.3	1786	1	US-07-920-430-19 Sequence 19, Appl
	769	16	0.3	1786	1	US-08-066-299-9 Sequence 9, Appli
	770	16	0.3	1786	1	US-08-265-047-1 Sequence 1, Appli
	771	16	0.3	1786	1	US-08-251-464-19 Sequence 19, Appl

	772	16	0.3	1786	4	PCT-US92-01364-19 Sequence 19, Appl
	773	16	0.3	1787	3	US-08-923-454A-3 Sequence 3, Appli
	774	16	0.3	1794	3	US-09-213-719-1 Sequence 1, Appli
	775	16	0.3	1804	2	US-08-504-459-5 Sequence 5, Appli
	776	16	0.3	1811	2	US-08-808-931-9 Sequence 9, Appli
	777	16	0.3	1811	3	US-08-808-323-9 Sequence 9, Appli
	778	16	0.3	1811	3	US-09-050-603A-9 Sequence 9, Appli
	779	16	0.3	1811	3	US-09-102-420B-9 Sequence 9, Appli
	780	16	0.3	1828	3	US-08-362-525-11 Sequence 11, Appl
	781	16	0.3	1847	2	US-08-808-931-11 Sequence 11, Appl
	782	16	0.3	1847	3	US-08-808-323-11 Sequence 11, Appl
	783	16	0.3	1847	3	US-09-050-603A-11 Sequence 11, Appl
	784	16	0.3	1847	3	US-09-102-420B-11 Sequence 11, Appl
	785	16	0.3	1849	1	US-08-399-696-3 Sequence 3, Appli
	786	16	0.3	1859	1	US-08-293-563-1 Sequence 1, Appli
	787	16	0.3	1866	3	US-08-909-742-1 Sequence 1, Appli
	788	16	0.3	1868	1	US-08-392-367B-1 Sequence 1, Appli
	789	16	0.3	1868	3	US-08-893-467A-1 Sequence 1, Appli
c	790	16	0.3	1897	1	US-08-184-632-1 Sequence 1, Appli
	791	16	0.3	1906	1	US-08-207-904-18 Sequence 18, Appl
	792	16	0.3	1927	2	US-09-040-799-1 Sequence 1, Appli
	793	16	0.3	1941	2	US-09-008-960-2 Sequence 2, Appli
	794	16	0.3	1941	3	US-09-368-240-2 Sequence 2, Appli
	795	16	0.3	1949	2	US-08-937-466-3 Sequence 3, Appli
	796	16	0.3	1949	2	US-09-172-528-3 Sequence 3, Appli
	797	16	0.3	1949	3	US-09-318-199-3 Sequence 3, Appli
	798	16	0.3	1954	1	US-08-050-132A-3 Sequence 3, Appli
	799	16	0.3	1954	2	US-07-989-847-3 Sequence 3, Appli
	800	16	0.3	1954	3	US-08-750-222A-3 Sequence 3, Appli
	801	16	0.3	1954	3	US-08-815-652B-3 Sequence 3, Appli
	802	16	0.3	1954	3	US-07-721-847A-5 Sequence 5, Appli
	803	16	0.3	1954	4	PCT-US92-05374A-3 Sequence 3, Appli
	804	16	0.3	1954	4	PCT-US95-07084-3 Sequence 3, Appli
	805	16	0.3	1954	5	5166058-5 Patent No. 5166058
	806	16	0.3	1962	1	US-08-399-696-1 Sequence 1, Appli
	807	16	0.3	1968	1	US-08-321-071A-17 Sequence 17, Appl
	808	16	0.3	1969	1	US-07-937-609-28 Sequence 28, Appl
c	809	16	0.3	1969	2	US-08-346-429-7 Sequence 7, Appli
c	810	16	0.3	1969	3	US-08-966-318-4 Sequence 4, Appli
	811	16	0.3	1972	1	US-08-463-048-1 Sequence 1, Appli
	812	16	0.3	1972	1	US-08-463-229-1 Sequence 1, Appli
	813	16	0.3	1972	2	US-08-302-891-1 Sequence 1, Appli
	814	16	0.3	2000	3	US-09-039-555B-19 Sequence 19, Appl
	815	16	0.3	2005	1	US-07-916-901-5 Sequence 5, Appli
c	816	16	0.3	2006	1	US-08-242-098-36 Sequence 36, Appl
	817	16	0.3	2022	2	US-08-937-540-7 Sequence 7, Appli
c	818	16	0.3	2040	2	US-08-031-538-10 Sequence 10, Appl
	819	16	0.3	2040	3	US-08-581-148C-21 Sequence 21, Appl
c	820	16	0.3	2058	2	US-08-749-391-1 Sequence 1, Appli
c	821	16	0.3	2058	3	US-09-390-200-1 Sequence 1, Appli
	822	16	0.3	2090	2	US-08-439-814-1 Sequence 1, Appli

	823	16	0.3	2096	3	US-09-008-481A-10 Sequence 10, Appl
	824	16	0.3	2096	3	US-09-195-666A-16 Sequence 16, Appl
	825	16	0.3	2104	1	US-08-592-126-96 Sequence 96, Appl
c	826	16	0.3	2127	1	US-08-832-883-54 Sequence 54, Appl
c	827	16	0.3	2127	2	US-08-832-877-54 Sequence 54, Appl
	828	16	0.3	2152	1	US-07-779-890-1 Sequence 1, Appli
	829	16	0.3	2152	1	US-07-779-890-1 Sequence 1, Appli
	830	16	0.3	2152	4	PCT-US93-05640-1 Sequence 1, Appli
	831	16	0.3	2205	3	US-08-888-077A-41 Sequence 41, Appl
c	832	16	0.3	2213	1	US-08-285-641-20 Sequence 20, Appl
	833	16	0.3	2245	1	US-08-203-905B-1 Sequence 1, Appli
	834	16	0.3	2248	1	US-08-639-217-1 Sequence 1, Appli
	835	16	0.3	2248	1	US-08-285-641-16 Sequence 16, Appl
	836	16	0.3	2248	1	US-08-975-405-1 Sequence 1, Appli
	837	16	0.3	2250	3	US-08-763-938-1 Sequence 1, Appli
	838	16	0.3	2254	1	US-08-153-848-27 Sequence 27, Appl
	839	16	0.3	2254	3	US-09-299-843A-27 Sequence 27, Appl
	840	16	0.3	2254	4	PCT-US93-11153-27 Sequence 27, Appl
	841	16	0.3	2258	1	US-07-720-589-1 Sequence 1, Appli
	842	16	0.3	2258	2	US-08-785-190-1 Sequence 1, Appli
	843	16	0.3	2258	4	PCT-US92-05539-1 Sequence 1, Appli
	844	16	0.3	2261	1	US-08-272-892D-1 Sequence 1, Appli
	845	16	0.3	2277	3	US-09-358-685-1 Sequence 1, Appli
	846	16	0.3	2295	2	US-08-842-842-6 Sequence 6, Appli
	847	16	0.3	2298	3	US-08-772-440-1 Sequence 1, Appli
	848	16	0.3	2303	3	US-08-836-567-1 Sequence 1, Appli
	849	16	0.3	2318	3	US-09-426-072-1 Sequence 1, Appli
	850	16	0.3	2325	2	US-08-966-388-5 Sequence 5, Appli
	851	16	0.3	2325	3	US-09-188-403-5 Sequence 5, Appli
	852	16	0.3	2325	3	US-09-188-404-5 Sequence 5, Appli
	853	16	0.3	2360	1	US-08-145-681-1 Sequence 1, Appli
	854	16	0.3	2360	1	US-08-250-308-1 Sequence 1, Appli
	855	16	0.3	2360	1	US-08-453-703-1 Sequence 1, Appli
	856	16	0.3	2360	2	US-08-456-106-1 Sequence 1, Appli
	857	16	0.3	2360	3	US-08-456-108-1 Sequence 1, Appli
	858	16	0.3	2360	4	PCT-US93-03614-1 Sequence 1, Appli
	859	16	0.3	2377	2	US-08-967-101-26 Sequence 26, Appl
	860	16	0.3	2377	2	US-08-592-541-26 Sequence 26, Appl
	861	16	0.3	2377	3	US-09-124-698-26 Sequence 26, Appl
c	862	16	0.3	2381	2	US-08-318-826A-9 Sequence 9, Appli
c	863	16	0.3	2395	3	US-08-985-335-6 Sequence 6, Appli
c	864	16	0.3	2400	5	5215909-13 Patent No. 5215909
c	865	16	0.3	2416	2	US-08-318-826A-8 Sequence 8, Appli
	866	16	0.3	2422	1	US-07-867-106-5 Sequence 5, Appli
c	867	16	0.3	2422	1	US-07-867-106-5 Sequence 5, Appli
	868	16	0.3	2444	3	US-08-906-791-1 Sequence 1, Appli
c	869	16	0.3	2445	5	5215909-9 Patent No. 5215909
	870	16	0.3	2452	1	US-07-756-250-15 Sequence 15, Appl
	871	16	0.3	2454	3	US-08-890-615-1 Sequence 1, Appli
	872	16	0.3	2499	3	US-08-699-103B-2 Sequence 2, Appli
	873	16	0.3	2522	3	US-09-062-416-13 Sequence 13, Appl

	874	16	0.3	2539	3	US-08-581-148C-20 Sequence 20, Appl
c	875	16	0.3	2549	1	US-08-470-720-2 Sequence 2, Appli
	876	16	0.3	2551	3	US-08-923-454A-26 Sequence 26, Appl
	877	16	0.3	2598	3	US-08-745-892-20 Sequence 20, Appl
	878	16	0.3	2599	2	US-08-239-276-7 Sequence 7, Appli
	879	16	0.3	2599	2	US-08-468-579B-7 Sequence 7, Appli
	880	16	0.3	2599	3	US-08-468-577B-7 Sequence 7, Appli
c	881	16	0.3	2624	1	US-08-232-144-3 Sequence 3, Appli
c	882	16	0.3	2651	1	US-08-462-949-1 Sequence 1, Appli
c	883	16	0.3	2651	1	US-08-023-764B-1 Sequence 1, Appli
c	884	16	0.3	2656	2	US-08-685-625A-5 Sequence 5, Appli
	885	16	0.3	2664	2	US-08-942-819-1 Sequence 1, Appli
	886	16	0.3	2672	3	US-09-214-564A-5 Sequence 5, Appli
	887	16	0.3	2720	2	US-08-007-107-3 Sequence 3, Appli
	888	16	0.3	2744	3	US-09-071-101-1 Sequence 1, Appli
	889	16	0.3	2744	3	US-09-369-618-1 Sequence 1, Appli
	890	16	0.3	2744	3	US-09-369-617-1 Sequence 1, Appli
c	891	16	0.3	2760	1	US-08-101-593-1 Sequence 1, Appli
c	892	16	0.3	2760	1	US-08-101-593-3 Sequence 3, Appli
	893	16	0.3	2763	1	US-08-248-466B-2 Sequence 2, Appli
	894	16	0.3	2782	2	US-08-937-466-1 Sequence 1, Appli
	895	16	0.3	2782	2	US-09-172-528-1 Sequence 1, Appli
	896	16	0.3	2782	3	US-09-318-199-1 Sequence 1, Appli
	897	16	0.3	2815	3	US-09-214-564A-1 Sequence 1, Appli
	898	16	0.3	2862	4	PCT-US96-12860-13 Sequence 13, Appl
	899	16	0.3	2900	3	US-08-038-832-1 Sequence 1, Appli
	900	16	0.3	2944	2	US-08-696-944-18 Sequence 18, Appl
	901	16	0.3	2953	2	US-08-859-201-1 Sequence 1, Appli
	902	16	0.3	2965	3	US-08-906-360-2 Sequence 2, Appli
c	903	16	0.3	2986	3	US-09-062-416-1 Sequence 1, Appli
	904	16	0.3	2989	1	US-08-687-379-9 Sequence 9, Appli
	905	16	0.3	2989	1	US-08-687-379-11 Sequence 11, Appl
c	906	16	0.3	3016	2	US-08-344-155C-97 Sequence 97, Appl
	907	16	0.3	3018	1	US-08-347-718B-3 Sequence 3, Appli
	908	16	0.3	3018	1	US-08-482-262-3 Sequence 3, Appli
	909	16	0.3	3018	5	5200183-1 Patent No. 5200183
c	910	16	0.3	3024	5	5284931-1 Patent No. 5284931
	911	16	0.3	3083	1	US-07-718-575-5 Sequence 5, Appli
	912	16	0.3	3083	1	US-08-481-206-5 Sequence 5, Appli
	913	16	0.3	3083	2	US-08-486-269A-5 Sequence 5, Appli
	914	16	0.3	3103	3	US-08-826-246-3 Sequence 3, Appli
	915	16	0.3	3103	3	US-08-944-495-3 Sequence 3, Appli
	916	16	0.3	3103	3	US-09-126-640-2 Sequence 2, Appli
	917	16	0.3	3111	3	US-09-487-444-3 Sequence 3, Appli
c	918	16	0.3	3138	1	US-07-867-106-4 Sequence 4, Appli
c	919	16	0.3	3138	3	US-09-234-332-5 Sequence 5, Appli
c	920	16	0.3	3155	2	US-08-591-629-7 Sequence 7, Appli
	921	16	0.3	3157	5	5198347-3 Patent No. 5198347
	922	16	0.3	3169	2	US-08-588-258B-27 Sequence 27, Appl
	923	16	0.3	3169	3	US-08-460-505-27 Sequence 27, Appl
	924	16	0.3	3169	4	PCT-US96-08295-27 Sequence 27, Appl

c	925	16	0.3	3189	3	US-08-875-223-4 Sequence 4, Appli
	926	16	0.3	3224	3	US-09-079-415-3 Sequence 3, Appli
	927	16	0.3	3250	3	US-08-617-860B-1 Sequence 1, Appli
	928	16	0.3	3294	1	US-08-409-995-1 Sequence 1, Appli
	929	16	0.3	3294	3	US-08-685-467-1 Sequence 1, Appli
	930	16	0.3	3348	1	US-08-222-616-34 Sequence 34, Appl
	931	16	0.3	3348	4	PCT-US95-04228-34 Sequence 34, Appl
c	932	16	0.3	3377	5	5198345-16 Patent No. 5198345
	933	16	0.3	3394	1	US-08-159-784-4 Sequence 4, Appli
c	934	16	0.3	3397	5	5210189-3 Patent No. 5210189
	935	16	0.3	3417	2	US-08-978-458-7 Sequence 7, Appli
	936	16	0.3	3417	3	US-08-978-454-7 Sequence 7, Appli
	937	16	0.3	3441	2	US-08-742-753-1 Sequence 1, Appli
	938	16	0.3	3467	3	US-08-972-719-1 Sequence 1, Appli
	939	16	0.3	3513	1	US-07-828-788A-5 Sequence 5, Appli
	940	16	0.3	3513	1	US-08-278-685-1 Sequence 1, Appli
	941	16	0.3	3513	1	US-08-278-685-3 Sequence 3, Appli
	942	16	0.3	3513	1	US-08-277-721-1 Sequence 1, Appli
	943	16	0.3	3513	1	US-08-277-721-3 Sequence 3, Appli
	944	16	0.3	3513	1	US-08-602-737-3 Sequence 3, Appli
	945	16	0.3	3513	4	PCT-US92-11337-5 Sequence 5, Appli
	946	16	0.3	3520	2	US-08-933-750C-94 Sequence 94, Appl
	947	16	0.3	3520	3	US-09-234-613-94 Sequence 94, Appl
	948	16	0.3	3522	1	US-08-040-751-4 Sequence 4, Appli
	949	16	0.3	3522	1	US-08-291-368-1 Sequence 1, Appli
	950	16	0.3	3522	2	US-08-962-190-1 Sequence 1, Appli
	951	16	0.3	3522	4	PCT-US95-10310-1 Sequence 1, Appli
	952	16	0.3	3522	5	5164180-3 Patent No. 5164180
c	953	16	0.3	3535	2	US-08-618-408B-1 Sequence 1, Appli
	954	16	0.3	3558	1	US-08-602-737-5 Sequence 5, Appli
	955	16	0.3	3641	1	US-08-030-096-5 Sequence 5, Appli
	956	16	0.3	3810	2	US-08-475-844-8 Sequence 8, Appli
	957	16	0.3	3810	4	PCT-US95-08429-8 Sequence 8, Appli
	958	16	0.3	3820	3	US-08-990-140-3 Sequence 3, Appli
	959	16	0.3	3829	2	US-08-631-097-8 Sequence 8, Appli
	960	16	0.3	3829	3	US-08-810-712-6 Sequence 6, Appli
	961	16	0.3	3855	3	US-08-974-549A-4 Sequence 4, Appli
	962	16	0.3	3871	2	US-08-599-455B-3 Sequence 3, Appli
	963	16	0.3	3956	2	US-08-966-388-9 Sequence 9, Appli
	964	16	0.3	3956	3	US-09-188-403-9 Sequence 9, Appli
	965	16	0.3	3956	3	US-09-188-404-9 Sequence 9, Appli
	966	16	0.3	4003	1	US-08-257-999-1 Sequence 1, Appli
	967	16	0.3	4029	3	US-08-851-843A-173 Sequence 173, App
	968	16	0.3	4029	3	US-08-974-549A-292 Sequence 292, App
	969	16	0.3	4038	3	US-08-969-125-8 Sequence 8, Appli
	970	16	0.3	4168	3	US-08-836-567-11 Sequence 11, Appl
	971	16	0.3	4177	3	US-09-023-082A-23 Sequence 23, Appl
c	972	16	0.3	4571	1	US-08-232-144-5 Sequence 5, Appli
	973	16	0.3	4669	2	US-08-583-276-18 Sequence 18, Appl
	974	16	0.3	4669	2	US-08-752-447-1 Sequence 1, Appli
	975	16	0.3	4669	5	5206352-3 Patent No. 5206352

c 976	16	0.3	4699	1	US-08-487-890A-1	Sequence 1, Appli
c 977	16	0.3	4699	2	US-08-478-435-1	Sequence 1, Appli
c 978	16	0.3	4699	2	US-08-337-483-1	Sequence 1, Appli
c 979	16	0.3	4699	2	US-08-478-373-1	Sequence 1, Appli
c 980	16	0.3	4699	3	US-08-474-671-1	Sequence 1, Appli
c 981	16	0.3	4699	3	US-08-483-577A-1	Sequence 1, Appli
982	16	0.3	4705	1	US-07-998-003A-96	Sequence 96, Appl
983	16	0.3	4705	1	US-08-453-274B-96	Sequence 96, Appl
984	16	0.3	4705	1	US-08-453-695A-96	Sequence 96, Appl
985	16	0.3	4705	1	US-08-268-161A-96	Sequence 96, Appl
986	16	0.3	4705	2	US-08-453-702A-96	Sequence 96, Appl
987	16	0.3	4705	4	PCT-US93-12588-96	Sequence 96, Appl
988	16	0.3	4705	4	PCT-US95-08071-96	Sequence 96, Appl
989	16	0.3	4895	3	US-09-053-866-1	Sequence 1, Appli
990	16	0.3	4949	3	US-09-138-024-22	Sequence 22, Appl
c 991	16	0.3	4964	1	US-08-470-720-5	Sequence 5, Appli
c 992	16	0.3	5345	1	US-08-044-618-7	Sequence 7, Appli
993	16	0.3	5532	2	US-08-475-035-3	Sequence 3, Appli
994	16	0.3	5852	1	US-07-867-106-2	Sequence 2, Appli
c 995	16	0.3	5852	1	US-07-867-106-2	Sequence 2, Appli
996	16	0.3	5893	1	US-08-592-126-54	Sequence 54, Appl
997	16	0.3	5893	2	US-08-687-080-44	Sequence 44, Appl
c 998	16	0.3	5914	1	US-08-480-784-1	Sequence 1, Appli
c 999	16	0.3	5914	1	US-08-483-553-1	Sequence 1, Appli
c1000	16	0.3	5914	1	US-08-487-002-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-470-260-4
; Sequence 4, Application US/08470260
; Patent No. 6077706
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    APPLICANT:  Bugnoli, Massimo
;    APPLICANT:  Telford, John
;    APPLICANT:  Macchia, Giovanni
;    APPLICANT:  Rappuoli, Rino
;    TITLE OF INVENTION:  Helicobacter Pylori Proteins Useful
;    TITLE OF INVENTION:  for Vaccines and Diagnostics
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
```



```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,260
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/256,848
;      FILING DATE:  21-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0316.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2708
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5925 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-470-260-4

  Query Match           100.0%;  Score 5925;  DB 3;  Length 5925;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 5925;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccattttaagcaactccatagaccactaaagaaactttttttgaggctatctttgaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCCATTTTAAGCAACTCCATAGACCACTAAAGAAACTTTTTTTGAGGCTATCTTTGAAA 60

Qy     61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATCTGTCCTATTGATTTGTTTTCCATTTTGTTTCCCATGTGGATCTTGTGGATCACAAAC 120

Qy    121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaaacgctt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTTAATTATACATGCTATAGTAAGCATGACACACAAACCAAACTATTTTTAGAACGCTT 180

Qy    181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CATGTGCTCACCTTGACTAACCATTTCTCCAACCATACTTTAGCGTTGCATTTGATTTCT 240

Qy    241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttagcaaatttt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCAAAAAGATTCATTTCTTATTTCTTGTTCTTATTAAAGTTCTTTCATTTAGCAAATTT 300
```



```
Qy    301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTGTTAATTGTGGGTAAAAATGTGAATCGTCCTAGCCTTTAGACGCCTGCAACGATCGGG 360

Qy    361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaatgcttgatattgttgtat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTTTTTTCAATATTAATAATGATTAATGAAAAAAAAAAAAAATGCTTGATATTGTTGTAT 420

Qy    421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATGAGAATGTTCAAAGACATGAATTGACTACTCAAGCGTGTAGCGATTTTTAGCAGTCT 480

Qy    481 ttgacactaacaagataccgataggtatgaaactaggtatagtaaggagaaacaatgact 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTGACACTAACAAGATACCGATAGGTATGAAACTAGGTATAGTAAGGAGAAACAATGACT 540

Qy    541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaaccccgcagcaatttatc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AACGAAACCATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATC 600

Qy    601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGAT 660

Qy    661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAAGGAATCTCG 720

Qy    721 caatttaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagattcagtat 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGATTCAGTAT 780

Qy    781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTA 840

Qy    841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGT 900

Qy    901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATG 960

Qy    961 gaaaatatcatacaacccccctatccttgatgataaagagaaagcggagttttgaaatct 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GAAAATATCATACAACCCCCTATCCTTGATGATAAAGAGAAAGCGGAGTTTTTGAAATCT 1020

Qy   1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   1021 GCCAAACAATCTTTTGCAGGAATCATTATAGGGAATCAAATCCGAACGGATCAAAAGTTC 1080

Qy   1081 atgggcgtgttttgatgagtccttgaaagaaaggcaagaagcagaaaaaaaatggagagcct 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ATGGGCGTGTTTTGATGAGTCCTTGAAAGAAAGGCAAGAAGCAGAAAAAAATGGAGAGCCT 1140

Qy   1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACTGGTGGGGATTGGTTGGATATTTTTCTCTCATTTATATTTGACAAAAAACAATCTTCT 1200

Qy   1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GATGTCAAAGAAGCAATCAATCAAGAACCAGTTCCCCATGTCCAACCAGATATAGCCACT 1260

Qy   1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ACCACCACCGACATACAAGGCTTACCGCCTGAAGCTAGAGATTTACTTGATGAAAGGGGT 1320

Qy   1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AATTTTTCTAAATTCACTCTTGGCGATATGGAAATGTTAGATGTTGAGGGAGTCGCTGAC 1380

Qy   1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ATTGATCCCAATTACAAGTTCAATCAATTATTGATTCACAATAACGCTCTGTCTTCTGTG 1440

Qy   1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TTAATGGGGAGTCATAATGGCATAGAACCTGAAAAAGTTTCATTGTTGTATGGGGGCAAT 1500

Qy   1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GGTGGTCCTGGAGCTAGGCATGATTGGAACGCCACCGTTGGTTATAAAGACCAACAAGGC 1560

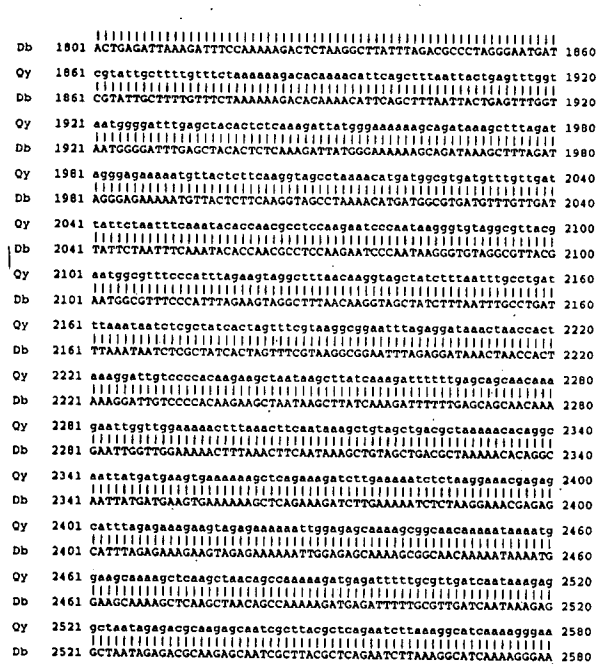
Qy   1561 aacaatgtggctacaataatttaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 1620

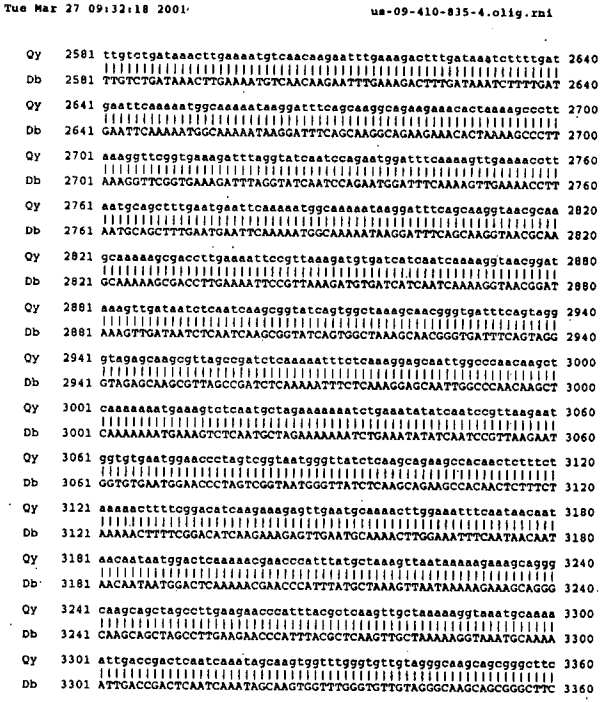
Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 1680

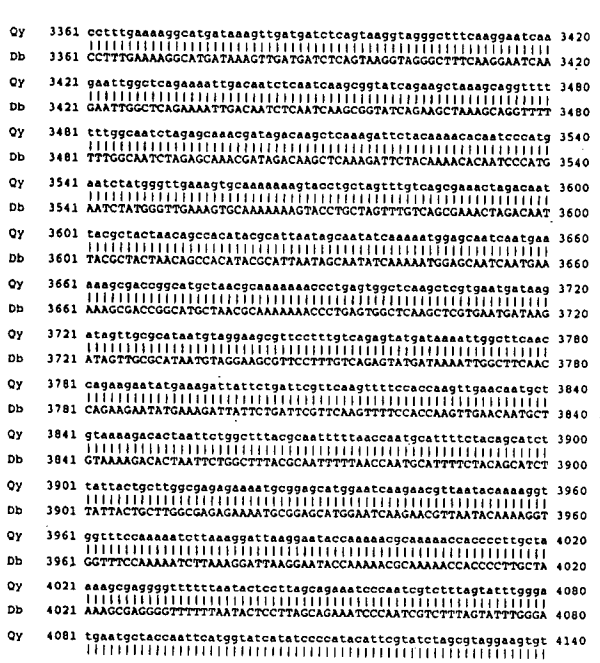
Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaatttt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 1740

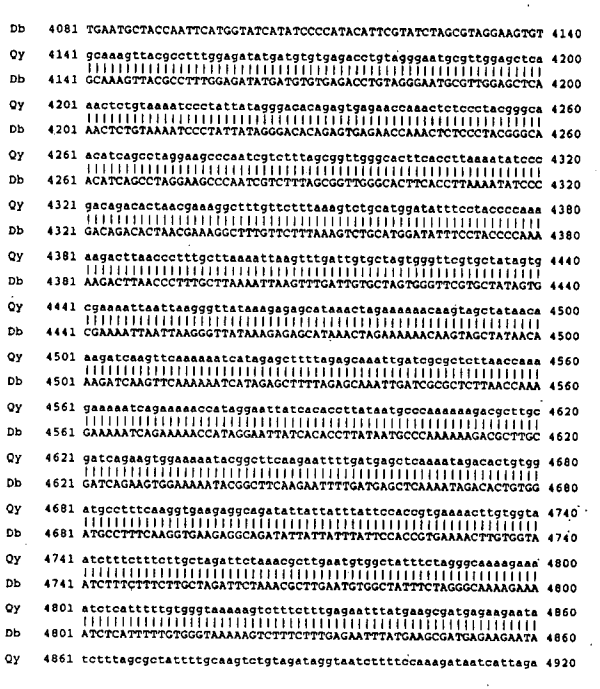
Qy   1741 cttgcacaaaataatgctaaatttagacaacttgagcgagaaagagaaggaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCGAGAAAGAGAAGGAAAAATTCCGA 1800

Qy   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
```









```
Db  4861 TCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 4920

Qy  4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacacaat 4980

Db  4921 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACACAAT 4980

Qy  4981 acccttatagtgtgagctatagccccctttttgggaattgagttatttttgactttaaaattt 5040

Db  4981 ACCCTTATAGTGTGAGCTATAGCCCCTTTTTTGGGAATTGAGTTATTTTGACTTTAAATTT 5040

Qy  5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100

Db  5041 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCC 5100

Qy  5101 gctcgcatgaaatttccactttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160

Db  5101 GCTCGCATGAAATTTCCACTTTAGGGAATGCGTGTGCATTTTTTTTTAAGGGCGTATTTTTG 5160

Qy  5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtagggggggtgttgaaagggc 5220

Db  5161 CTGCAAATATCCTACAATAGCATCGCCCGAATGGATGAGTAGGGGGGGTGTTGAAAGGGC 5220

Qy  5221 aaaatgctccataaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280

Db  5221 AAAATGCTCCATAAAATAGCCCTCAATTTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 5280

Qy  5281 taaaatcaatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340

Db  5281 TAAAATAATCACTTCGGGAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTTTCT 5340

Qy  5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400

Db  5341 AACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAAGAGAAGTGGCTAAA 5400

Qy  5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460

Db  5401 TGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGGATC 5460

Qy  5461 gtcgcttttgtccctagcactaaaatagggggcgtttttatctttacttgtcgcttgatc 5520

Db  5461 GTCGCTTTTGTCCCTAGCACTAAAATAGGGGCGTTTTTATCTTTACTTGTCGCTTGATC 5520

Qy  5521 gctaaaatgcttggctcaatcacgcccacaataggggattttggaatgcttttgcatctct 5580

Db  5521 GCTAAAATGCTTGGCTCAATCACGCCCACAATAGGGATTTTGGAATGCTTTTGCATCTCT 5580

Qy  5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640

Db  5581 TCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGGT 5640
```



```
Qy  5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700

Db  5641 TTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTGCCA 5700

Qy  5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760

Db  5701 TAAGGCACTCTAGCCGTATCGCCATAATAGATGATTTCATCAAATAATTGCGCTTTTAAA 5760

Qy  5761 aggctttttaaaacgctaaacccctcccacaccgcgtatcaaaaacgcctattttcatgaca 5820

Db  5761 AGGCTTTTTAAAACGCTAAACCCTCCCACACCGCTATCAAAAACGCCTATTTTCATGACA 5820

Qy  5821 ctttttaatttaatgggattaattagggatttttattttttcattcattaagtttaaaaat 5880

Db  5821 CTTTTTTAATTTAATGGGATTAATTAGGGATTTTATTTTTCATTCATTAAGTTTAAAAAT 5880

Qy  5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925

Db  5881 TCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATAGACAAAGCTT 5925


RESULT  2
US-08-471-491-4
; Sequence 4, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0044
;  CURRENT APPLICATION NUMBER: US/08/471,491B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;  LENGTH: 5925
;  TYPE: DNA
;  ORGANISM: Helicobacter pylori
US-08-471-491-4


  Query Match           100.0%;  Score 5925;  DB 3;  Length 5925;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 5925;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctccattttaagcaactccatagaccactaaagaaaccttttttggaggctatctttgaaa 60
```



```
Db       1 otccattttaagcaactccatagaccactaaagaaaccttttttggaggctatctttgaaa 60

Qy      61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaac 120

Db      61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaac 120

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaacccaaactatttttagaacgctt 180

Db     121 gcttaattatacatgctatagtaagcatgacacacaaacccaaactatttttagaacgctt 180

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240

Db     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240

Qy     241 tcaaaaaagattcatttcttatttcttgttcttattaaagttctttcatttagcaaattt 300

Db     241 tcaaaaaagattcatttcttatttcttgttcttattaaagttctttcatttagcaaattt 300

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360

Db     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360

Qy     361 ctttttttcaatattaataatgattaatgaaaaaaaaaaaaaaatgcttgatattgttgtat 420

Db     361 ctttttttcaatattaataatgattaatgaaaaaaaaaaaaaatgcttgatattgttgtat 420

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480

Db     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480

Qy     481 ttgacactaacaagatacgataggtatgaaactaggtatagtaaggagaaacaatgact 540

Db     481 ttgacactaacaagatacgataggtatgaaactaggtatagtaaggagaaacaatgact 540

Qy     541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600

Db     541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600

Qy     601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660

Db     601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660

Qy     661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720

Db     661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720

Qy     721 caatttaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780

Db     721 caatttaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780

Qy     781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
```



```
Db     781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840

Qy     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatttttcacaagt 900

Db     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatttttcacaagt 900

Qy     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttatg 960

Db     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttatg 960

Qy     961 gaaaatatcatacaacccccatcctttgatgataaagagaaagcggagttttgaaatct 1020

Db     961 gaaaatatcatacaacccccatcctttgatgataaagagaaagcggagttttgaaatct 1020

Qy    1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080

Db    1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080

Qy    1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140

Db    1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140

Qy    1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaaacaatcttct 1200

Db    1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaaacaatcttct 1200

Qy    1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260

Db    1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260

Qy    1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320

Db    1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320

Qy    1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgaggggagtcgctgac 1380

Db    1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgaggggagtcgctgac 1380

Qy    1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440

Db    1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440

Qy    1441 ttaatgggggagtcataatggcatagaaccctgaaaaagtttcattgttgtatgggggcaat 1500

Db    1441 ttaatgggggagtcataatggcatagaaccctgaaaaagtttcattgttgtatgggggcaat 1500

Qy    1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560

Db    1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
```

```
Qy   1561 aacaatgtggctacaatattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 aacaatgtggctacaatattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740

Qy   1741 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800

Qy   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860

Qy   1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920

Qy   1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980

Qy   1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040

Qy   2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100

Qy   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160

Qy   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220

Qy   2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280

Qy   2281 gaattggttggaaaaacttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 gaattggttggaaaaacttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
```



```
Qy   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400

Qy   2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460

Qy   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520

Qy   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580

Qy   2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640

Qy   2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagccctt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagccctt 2700

Qy   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaacctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaacctt 2760

Qy   2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820

Qy   2821 gcaaaaagcgacccttgaaaatttccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 gcaaaaagcgacccttgaaaatttccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880

Qy   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940

Qy   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaacaagct 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaacaagct 3000

Qy   3001 caaaaaaatgaaagtctcaatgctagaaaaaaaatctgaaatatatcaatccgttaagaat 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 caaaaaaatgaaagtctcaatgctagaaaaaaaatctgaaatatatcaatccgttaagaat 3060

Qy   3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120

Qy   3121 aaaaacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 aaaaacttttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180

Qy   3181 aacaatgaatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 aacaatgaatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300

Qy   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360

Qy   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420

Qy   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600

Qy   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720

Qy   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840

Qy   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcattttctacagcatct 3900
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcattttctacagcatct 3900

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960

Qy   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020

Qy   4021 aaagcgagggggtttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 aaagcgagggggtttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080

Qy   4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140

Qy   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380

Qy   4381 aagacttaaccctttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 aagacttaaccctttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440

Qy   4441 cgaaaattaatttaagggttataaagagagcataaactagaaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 cgaaaattaatttaagggttataaagagagcataaactagaaaaaaacaagtagctataaca 4500

Qy   4501 aagatcaagttcaaaaaatcatagagcttttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 aagatcaagttcaaaaaatcatagagcttttagagcaaattgatcgcgctcttaaccaaa 4560

Qy   4561 gaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 gaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaaagacgcttgc 4620
```

```
Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
Db   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680

Qy   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
Db   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740

Qy   4741 atctttctttcttgctagatttctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
Db   4741 atctttctttcttgctagatttctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
Db   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860

Qy   4861 tctttagcgctatttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920
Db   4861 tctttagcgctatttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920

Qy   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
Db   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980

Qy   4981 acccttatagtgtgagctatagcccctttttgggaattgagttatttttgactttaaatttt 5040
Db   4981 acccttatagtgtgagctatagcccctttttgggaattgagttatttttgactttaaatttt 5040

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
Db   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100

Qy   5101 gctcgcatgaaattccacttttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
Db   5101 gctcgcatgaaattccacttttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggggtgttgaaagggc 5220
Db   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggggtgttgaaagggc 5220

Qy   5221 aaaatgctccatataaaatagccctcaatttttttgagcgattaagggaaaatgcgtgcaacc 5280
Db   5221 aaaatgctccatataaaatagccctcaatttttttgagcgattaagggaaaatgcgtgcaacc 5280

Qy   5281 taaaatatcacttcgggaaaatctttaagggagtgaaatataacgcatgcaagtttct 5340
Db   5281 taaaatatcacttcgggaaaatctttaagggagtgaaatataacgcatgcaagtttct 5340

Qy   5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
Db   5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
```



```
Qy   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
Db   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460

Qy   5461 gtcgcttttgtccctagcactaaaaataggggcgtttttatcttttacttgtcgcttgatc 5520
Db   5461 gtcgcttttgtccctagcactaaaaataggggcgtttttatcttttacttgtcgcttgatc 5520

Qy   5521 gctaaaatgcttggctcaatcacgcccacaatagggattttggaatgcttttgcatctct 5580
Db   5521 gctaaaatgcttggctcaatcacgcccacaatagggattttggaatgcttttgcatctct 5580

Qy   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
Db   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640

Qy   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
Db   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700

Qy   5701 taaggcactctagccgtatcgccatagatgatttcatcaaataattgcgcttttaaa 5760
Db   5701 taaggcactctagccgtatcgccatagatgatttcatcaaataattgcgcttttaaa 5760

Qy   5761 agggtttttaaaacgctaaacccctcccacaccgctatcaaaaacgcctattttcatgaca 5820
Db   5761 agggtttttaaaacgctaaacccctcccacaccgctatcaaaaacgcctattttcatgaca 5820

Qy   5821 ctttttaatttaatgggattaattagggatttttattttttcattcattaagtttaaaaat 5880
Db   5821 ctttttaatttaatgggattaattagggatttttattttttcattcattaagtttaaaaat 5880

Qy   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
Db   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925

RESULT   3
US-08-466-662-4
; Sequence 4, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0057
```



```
;  CURRENT APPLICATION NUMBER: US/08/466,662B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 5925
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
US-08-466-662-4

  Query Match           100.0%;  Score 5925;  DB 3;  Length 5925;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 5925;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctccatttttaagcaactccatagaccactaaagaaacttttttgaggctatctttgaaa 60
Db       1 ctccatttttaagcaactccatagaccactaaagaaacttttttgaggctatctttgaaa 60

Qy      61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaaac 120
Db      61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaaac 120

Qy     121 gcttaattatatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180
Db     121 gcttaattatatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240
Db     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240

Qy     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttttagcaaattt 300
Db     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttttagcaaattt 300

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360
Db     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360

Qy     361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaaatgcttgatattgttgtat 420
Db     361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaatgcttgatattgttgtat 420

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
Db     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480

Qy     481 ttgacactaacaagataccgataggtatgaaactaggtatagtaaggagaaacaatgact 540
Db     481 ttgacactaacaagataccgataggtatgaaactaggtatagtaaggagaaacaatgact 540
```



```
Qy      541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600
Db      541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600

Qy      601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
Db      601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660

Qy      661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720
Db      661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720

Qy      721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780
Db      721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780

Qy      781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
Db      781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840

Qy      841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatttttcacaagt 900
Db      841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaatttttcacaagt 900

Qy      901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatg 960
Db      901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatg 960

Qy      961 gaaaatatcatacaaccccctatccttgatgataaagagaaagcggagttttgaaatct 1020
Db      961 gaaaatatcatacaaccccctatccttgatgataaagagaaagcggagttttgaaatct 1020

Qy     1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080
Db     1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080

Qy     1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140
Db     1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140

Qy     1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200
Db     1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200

Qy     1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260
Db     1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260

Qy     1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
Db     1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
```

```
Qy   1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380

Qy   1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440

Qy   1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500

Qy   1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaaagaccaacaaggc 1560

Qy   1561 aacaatgtggctacaatаattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 aacaatgtggctacaatаattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740

Qy   1741 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 cttgcacaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaaattccga 1800

Qy   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgcccctagggaatgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgcccctagggaatgat 1860

Qy   1861 cgtattgcttttgtttctaaaaaagacaccaaaacattcagctttaattactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 cgtattgcttttgtttctaaaaaagacaccaaaacattcagctttaattactgagtttggt 1920

Qy   1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980

Qy   1981 agggagaaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 agggagaaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040

Qy   2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100

Qy   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160

Qy   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220

Qy   2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280

Qy   2281 gaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 gaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340

Qy   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400

Qy   2401 catttagagaaagaagtagagaaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 catttagagaaagaagtagagaaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460

Qy   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520

Qy   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580

Qy   2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640

Qy   2641 gaattccaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagcccct 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 gaattccaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagcccct 2700

Qy   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaaccctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaaccctt 2760

Qy   2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820

Qy   2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880

Qy   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940

Qy   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagccaattggcccaacaagct 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagccaattggcccaacaagct 3000

Qy   3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060

Qy   3061 ggtgtgaatggaacccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ggtgtgaatggaacccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120

Qy   3121 aaaaactttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 aaaaactttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180

Qy   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300

Qy   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360

Qy   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420

Qy   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
```



```
Qy   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720

Qy   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840

Qy   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960

Qy   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccctgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccctgcta 4020

Qy   4021 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080

Qy   4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140

Qy   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
```

```
Qy   4381 aagacttaaccctttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 aagacttaaccctttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500

Qy   4501 aagatcaagttcaaaaaatcatagagctttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 aagatcaagttcaaaaaatcatagagctttagagcaaattgatcgcgctcttaaccaaa 4560

Qy   4561 gaaaaatcagaaaaaccatagaattatcacaccttataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 gaaaaatcagaaaaaccatagaattatcacaccttataatgcccaaaaaagacgcttgc 4620

Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680

Qy   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860

Qy   4861 tctttagcgctattttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 tctttagcgctattttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920

Qy   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980

Qy   4981 acccttatagtgtgagctatagccccttttttgggaattgagttatttttgactttaaattt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 acccttatagtgtgagctatagccccttttttgggaattgagttatttttgactttaaattt 5040

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100

Qy   5101 gctcgcatgaaattccactttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   5101 gctcgcatgaaattccactttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtagggggggtgttgaaagggc 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtagggggggtgttgaaagggc 5220

Qy   5221 aaaatgctccatataaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 aaaatgctccatataaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280

Qy   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340

Qy   5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400

Qy   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460

Qy   5461 gtcgcttttgtccctagcactaaaatagggggcgttttatctttacttgtcgcttgatc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 gtcgcttttgtccctagcactaaaatagggggcgttttatctttacttgtcgcttgatc 5520

Qy   5521 gctaaaaatgcttggctcaatcacgcccacaataggggattttggaatgcttttgcatctct 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 gctaaaaatgcttggctcaatcacgcccacaataggggattttggaatgcttttgcatctct 5580

Qy   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640

Qy   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700

Qy   5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760

Qy   5761 aggctttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctattttcatgaca 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 aggctttttaaaacgctaaaccctcccacaccgctatcaaaaacgcctattttcatgaca 5820

Qy   5821 ctttttaatttaatgggattaattagggatttattttttcattcattaagtttaaaaat 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 ctttttaatttaatgggattaattagggatttattttttcattcattaagtttaaaaat 5880

Qy   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
```



```
          ||||||||||||||||||||||||||||||||||||||||||||
Db   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
```

```
RESULT   4
US-08-477-451-25
; Sequence 25, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19932 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-25


  Query Match           100.0%;  Score 5925;  DB 2;  Length 19932;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 5925;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctccatttaagcaactccatagaccactaaagaaacttttttgaggctatctttgaaa 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   14008 CTCCATTTTAAGCAACTCCATAGACCACTAAAGAAACTTTTTTTGAGGCTATCTTTGAAA 14067

Qy      61 atctgtcctattgatttgttttccatttttgtttccccatgtggatcttgtggatcacaaac 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14068 ATCTGTCCTATTGATTTGTTTTCCATTTTGTTTCCCATGTGGATCTTGTGGATCACAAAC 14127

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14128 GCTTAATTATACATGCTATAGTAAGCATGACACACAAACCAAACTATTTTTAGAACGCTT 14187

Qy     181 catgtgctcaccttgactaaccatttctccaaccatacttttagcgttgcatttgatttct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14188 CATGTGCTCACCTTGACTAACCATTTCTCCAACCATACTTTTAGCGTTGCATTTGATTTCT 14247

Qy     241 tcaaaaagattcatttcttatttottgttcttattaaagttcttttcatttttagcaaattt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14248 TCAAAAAGATTCATTTCTTATTTCTTGTTCTTATTAAAGTTCTTTCATTTTAGCAAATTT 14307

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14308 TTGTTAATTGTGGGTAAAAATGTGAATCGTCCTAGCCTTTAGACGCCTGCAACGATCGGG 14367

Qy     361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaaatgcttgatattgttgtat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14368 CTTTTTTCAATATTAATAATGATTAATGAAAAAAAAAAAAAATGCTTGATATTGTTGTAT 14427

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14428 AATGAGAATGTTCAAAGACATGAATTGACTACTCAAGCGTGTAGCGATTTTTAGCAGTCT 14487

Qy     481 ttgacactaacaagatacgatagtatgaaactaggtatagtaaggagaaacaatgact 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14488 TTGACACTAACAAGATACCGATAGGTATGAAACTAGGTATAGTAAGGAGAAACAATGACT 14547

Qy     541 aacgaaaccattgaccaacaaccacaaaccgaagcggctttttaacccgcagcaatttatc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14548 AACGAAACCATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATC 14607

Qy     601 aataatcttcaagtagcttttcttaaagttgatcaacgctgtcgcttcatacgatcctgat 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14608 AATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGAT 14667

Qy     661 caaaaaccaatcgttgataagaacgatagggataacagggcaagcttttgaaggaatctcg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14668 CAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAAGGAATCTCG 14727

Qy     721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14728 CAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTAT 14787

Qy     781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
```

```
Db 14788 TTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTA 14847
Qy   841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 900
Db 14848 GAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGT 14907
Qy   901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttatg 960
Db 14908 TGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATG 14967
Qy   961 gaaaatatcatacaaccccctatccttgatgataaagagaaagcggagtttttgaaatct 1020
Db 14968 GAAAATATCATACAACCCCCTATCCTTGATGATAAAGAGAAAGCGGAGTTTTTGAAATCT 15027
Qy  1021 gccaaacaatcttttgcaggaatcattataggggaatcaaatccgaacggatcaaaagttc 1080
Db 15028 GCCAAACAATCTTTTGCAGGAATCATTATAGGGAATCAAATCCGAACGGATCAAAAGTTC 15087
Qy  1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagcct 1140
Db 15088 ATGGGCGTGTTTGATGAGTCCTTGAAAGAAAGGCAAGAAGCAGAAAAAAATGGAGAGCCT 15147
Qy  1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200
Db 15148 ACTGGTGGGGATTGGTTGGATATTTTTCTCTCATTTATATTTGACAAAAAACAATCTTCT 15207
Qy  1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260
Db 15208 GATGTCAAAGAAGCAATCAATCAAGAACCAGTTCCCCATGTCCAACCAGATATAGCCACT 15267
Qy  1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
Db 15268 ACCACCACCGACATACAAGGCTTACCGCCTGAAGCTAGAGATTTACTTGATGAAAGGGGT 15327
Qy  1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380
Db 15328 AATTTTTCTAAATTCACTCTTGGCGATATGGAAATGTTAGATGTTGAGGGAGTCGCTGAC 15387
Qy  1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtctttctgtg 1440
Db 15388 ATTGATCCCAATTACAAGTTCAATCAATTATTGATTCACAATAACGCTCTGTCTTCTGTG 15447
Qy  1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
Db 15448 TTAATGGGGAGTCATAATGGCATAGAACCTGAAAAAGTTTCATTGTTGTATGGGGGCAAT 15507
Qy  1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
Db 15508 GGTGGTCCTGGAGCTAGGCATGATTGGAACGCCACCGTTGGTTATAAAGACCAACAAGGC 15567
```



```
Qy  1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
Db 15568 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 15627
Qy  1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
Db 15628 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 15687
Qy  1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
Db 15688 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 15747
Qy  1741 cttgcacaaaaataatgctaaattagacaacttgagcgagaaagagaagggaaaaattccga 1800
Db 15748 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCGAGAAAGAGAAGGAAAAATTCCGA 15807
Qy  1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
Db 15808 ACTGAGATTAAAGATTTCCAAAAAGACTCTAAGGCTTATTTAGACGCCCTAGGGAATGAT 15867
Qy  1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
Db 15868 CGTATTGCTTTTGTTTCTAAAAAAGACACAAAACATTCAGCTTTAATTACTGAGTTTGGT 15927
Qy  1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagatagtaaagctttagat 1980
Db 15928 AATGGGGATTTGAGCTACACTCTCAAAGATTATGGGAAAAAAGCAGATAAAGCTTTAGAT 15987
Qy  1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
Db 15988 AGGGAGAAAAATGTTACTCTTCAAGGTAGCCTAAAACATGATGGCGTGATGTTTGTTGAT 16047
Qy  2041 tattctaatttcaaatacaccaacgcctccaagaatccaatagggtgtaggcgttacg 2100
Db 16048 TATTCTAATTTCAAATACACCAACGCCTCCAAGAATCCCAATAAGGGTGTAGGCGTTACG 16107
Qy  2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
Db 16108 AATGGCGTTTCCCATTTAGAAGTAGGCTTTAACAAGGTAGCTATCTTTAATTTGCCTGAT 16167
Qy  2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
Db 16168 TTAAATAATCTCGCTATCACTAGTTTCGTAAGGCGGAATTTAGAGGATAAACTAACCACT 16227
Qy  2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280
Db 16228 AAAGGATTGTCCCCACAAGAAGCTAATAAGCTTATCAAAGATTTTTGAGCAGCAACAAA 16287
Qy  2281 gaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
Db 16288 GAATTGGTTGGAAAAACTTTAAACTTCAATAAAGCTGTAGCTGACGCTAAAAACACAGGC 16347
```



```
Qy  2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
Db 16348 AATTATGATGAAGTGAAAAAAGCTCAGAAAGATCTTGAAAAATCTCTAAGGAAACGAGAG 16407
Qy  2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaataaaatg 2460
Db 16408 CATTTAGAGAAAGAAGTAGAGAAAAAATTGGAGAGCAAAAGCGGCAACAAAAATAAAATG 16467
Qy  2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
Db 16468 GAAGCAAAAGCTCAAGCTAACAGCCAAAAAGATGAGATTTTTGCGTTGATCAATAAAGAG 16527
Qy  2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
Db 16528 GCTAATAGAGACGCAAGAGCAATCGCTTACGCTCAGAATCTTAAAGGCATCAAAAGGGAA 16587
Qy  2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
Db 16588 TTGTCTGATAAACTTGAAAATGTCAACAAGAATTTGAAAGACTTTGATAAATCTTTTGAT 16647
Qy  2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagccctt 2700
Db 16648 GAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGCAGAAGAAACACTAAAAGCCCTT 16707
Qy  2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaacctt 2760
Db 16708 AAAGGTTCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAAAGTTGAAAACCTT 16767
Qy  2761 aatgcagctttgaatgaattcaaaaatggcaaaaatagggatttcagcaaggtaacgcaa 2820
Db 16768 AATGCAGCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGTAACGCAA 16827
Qy  2821 gcaaaaagcgacctttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
Db 16828 GCAAAAAGCGACCTTGAAAATTCCGTTAAAGATGTGATCATCAATCAAAAGGTAACGGAT 16887
Qy  2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
Db 16888 AAAGTTGATAATCTCAATCAAGCGGTATCAGTGGCTAAAGCAACGGGTGATTTCAGTAGG 16947
Qy  2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaacaagct 3000
Db 16948 GTAGAGCAAGCGTTAGCCGATCTCAAAAATTTCTCAAAGGAGCAATTGGCCCAACAAGCT 17007
Qy  3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060
Db 17008 CAAAAAAATGAAAGTCTCAATGCTAGAAAAAAATCTGAAATATATCAATCCGTTAAGAAT 17067
Qy  3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
```



```
Db 17068 GGTGTGAATGGAACCCTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACTCTTTCT 17127
Qy  3121 aaaaactttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
Db 17128 AAAAACTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAAT 17187
Qy  3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
Db 17188 AACAATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGG 17247
Qy  3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
Db 17248 CAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAA 17307
Qy  3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
Db 17308 ATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC 17367
Qy  3361 cotttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
Db 17368 CCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAA 17427
Qy  3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
Db 17428 GAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTT 17487
Qy  3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaaacacaatcccatg 3540
Db 17488 TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATG 17547
Qy  3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
Db 17548 AATCTATGGGTTGAAAGTGCAAAAAAAGTACCTGCTAGTTTGTCAGCGAAACTAGACAAT 17607
Qy  3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
Db 17608 TACGCTACTAACAGCCACATACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAA 17667
Qy  3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
Db 17668 AAAGCGACCGGCATGCTAACGCAAAAAAACCCTGAGTGGCTCAAGCTCGTGAATGATAAG 17727
Qy  3721 atagttgcgcataatgtaggaagcgttccttttgtcagagtatgataaaaattggcttcaac 3780
Db 17728 ATAGTTGCGCATAATGTAGGAAGCGTTCCTTTGTCAGAGTATGATAAAATTGGCTTCAAC 17787
Qy  3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
Db 17788 CAGAAGAATATGAAAGATTATTCTGATTCGTTCAAGTTTTCCACCAAGTTGAACAATGCT 17847
Qy  3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcattttctacagcatct 3900
```

```
Db  17848 GTAAAAGACACTAATTCTGGCTTTACGCAATTTTTAACCAATGCATTTTCTACAGCATCT 17907

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17908 TATTACTGCTTGGCGAGAGAAAATGCGGAGCATGGAATCAAGAACGTTAATACAAAAGGT 17967

Qy   3961 ggttttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17968 GGTTTCCAAAAATCTTAAAGGATTAAGGAATACCAAAAACGCAAAAACCACCCCTTGCTA 18027

Qy   4021 aaagcgaggggttttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18028 AAAGCGAGGGGTTTTTTAATACTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGA 18087

Qy   4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18088 TGAATGCTACCAATTCATGGTATCATATCCCCATACATTCGTATCTAGCGTAGGAAGTGT 18147

Qy   4141 gcaaagttacgcctttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18148 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 18207

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18208 AACTCTGTAAAATCCCTATTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 18267

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18268 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 18327

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18328 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAA 18387

Qy   4381 aagacttaacccttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18388 AAGACTTAACCCTTTGCTTAAAATTAAGTTTGATTGTGCTAGTGGGTTCGTGCTATAGTG 18447

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18448 CGAAAATTAATTAAGGGTTATAAAGAGAGCATAAACTAGAAAAAACAAGTAGCTATAACA 18507

Qy   4501 aagatcaagttcaaaaaatcatagagcttttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18508 AAGATCAAGTTCAAAAAATCATAGAGCTTTTAGAGCAAATTGATCGCGCTCTTAACCAAA 18567

Qy   4561 gaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18568 GAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAAGACGCTTGC 18627
```



```
Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18628 GATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACTGTGG 18687

Qy   4681 atgccttttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18688 ATGCCTTTCAAGGTGAAGAGGCAGATATTATTATTTATTCCACCGTGAAAACTTGTGGTA 18747

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18748 ATCTTTCTTTCTTGCTAGATTCTAAACGCTTGAATGTGGCTATTTCTAGGGCAAAAGAAA 18807

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18808 ATCTCATTTTTGTGGGTAAAAAGTCTTTCTTTGAGAATTTATGAAGCGATGAGAAGAATA 18867

Qy   4861 tctttagcgctatttttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18868 TCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 18927

Qy   4921 catttcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18928 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACACAAT 18987

Qy   4981 acccttatagtgtgagctatagccccttttttgggaattgagttatttttgacttttaaattt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18988 ACCCTTATAGTGTGAGCTATAGCCCCTTTTTGGGAATTGAGTTATTTTGACTTTAAATTT 19047

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19048 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCC 19107

Qy   5101 gctcgcatgaaattccactttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19108 GCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATTTTTTTTAAGGGCGTATTTTTG 19167

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtagggggggtgttgaaagggc 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19168 CTGCAAATATCCTACAATAGCATCGCCCGAATGGATGAGTAGGGGGGGTGTTGAAAGGGC 19227

Qy   5221 aaaatgctccataaaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19228 AAAATGCTCCATAAAATAGCCCTCAATTTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 19287

Qy   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19288 TAAAATAATCACTTCGGGAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTTTCT 19347

Qy   5341 aacaattcgcccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19348 AACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAAGAGAAGTGGCTAAA 19407
```



```
Qy   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19408 TGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGGATC 19467

Qy   5461 gtcgcttttgtcccctagcactaaaataggggcgtttttatctttttacttgtcgcttgatc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19468 GTCGCTTTTGTCCCTAGCACTAAAATAGGGGCGTTTTTATCTTTACTTGTCGCTTGATC 19527

Qy   5521 gctaaaaatgcttggctcaatcacgcccacaatagggattttggaatgcttttgcatctct 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19528 GCTAAAATGCTTGGCTCAATCACGCCCACAATAGGGATTTTGGAATGCTTTTGCATCTCT 19587

Qy   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19588 TCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGGT 19647

Qy   5641 ttgaaaaaatccaaagcctctaagccaaatttgcttgatcgtagtggggtctttagtgcca 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19648 TTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTGCCA 19707

Qy   5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19708 TAAGGCACTCTAGCCGTATCGCCATAATAGATGATTTCATCAAATAATTGCGCTTTTAAA 19767

Qy   5761 aggctttttaaaacgctaaacccctcccacaccgctatcaaaaacgcctattttcatgaca 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19768 AGGCTTTTTAAAACGCTAAACCCTCCCACACCGCTATCAAAAACGCCTATTTTCATGACA 19827

Qy   5821 ctttttaatttaatgggattaattagggatttattttttcattcattaagtttaaaaat 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19828 CTTTTTAATTTAATGGGATTAATTAGGGATTTATTTTTCATTCATTAAGTTTAAAAAT 19887

Qy   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
          |||||||||||||||||||||||||||||||||||||||||||||
Db  19888 TCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATAGACAAAGCTT 19932


RESULT   5
US-08-477-451-1
; Sequence 1, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
```



```
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10299 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-1


  Query Match             99.1%;  Score 5874;  DB 2;  Length 10299;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 5924;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ctccatttttaagcaactccatagaccactaaagaaacttttttttgaggctatctttgaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4375 CTCCATTTTAAGCAACTCCATAGACCACTAAAGAAACTTTTTTTGAGGCTATCTTTGAAA 4434

Qy      61 atctgtcctattgatttgttttccattttgtttcccatgtggatcttgtggatcacaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4435 ATCTGTCCTATTGATTTGTTTTCCATTTTGTTTCCCATGTGGATCTTGTGGATCACAAAC 4494

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4495 GCTTAATTATACATGCTATAGTAAGCATGACACACAAACCAAACTATTTTTAGAACGCTT 4554

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4555 CATGTGCTCACCTTGACTAACCATTTCTCCAACCATACTTTAGCGTTGCATTTGATTTCT 4614

Qy     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttttagcaaattt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4615 TCAAAAAGATTCATTTCTTATTTCTTGTTCTTATTAAAGTTCTTTCATTTTAGCAAATTT 4674
```

```
Qy    301 ttgttaattgtgggtaaaaatgtgaatcgtcctagcctttagacgcctgcaacgatcggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4675 TTGTTAATTGTGGGTAAAAATGTGAATCGTCCTAGCCTTTAGACGCCTGCAACGATCGGG 4734

Qy    361 ctttttttcaatattaataatgattaatgaaaaaaaaaaaaaaatgcttgatattgttgtat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4735 CTTTTTTCAATATTAATAATGATTAATGAAAAAAAAAAAAAAATGCTTGATATTGTTGTAT 4794

Qy    421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4795 AATGAGAATGTTCAAAGACATGAATTGACTACTCAAGCGTGTAGCGATTTTTAGCAGTCT 4854

Qy    481 ttgacactaacaagataccgataggtatgaaactaggtatagtaaggagaaacaatgact 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4855 TTGACACTAACAAGATACCGATAGGTATGAAACTAGGTATAGTAAGGAGAAACAATGACT 4914

Qy    541 aacgaaaccattgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4915 AACGAAACCATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATC 4974

Qy    601 aataatcttcaagtagctttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4975 AATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGAT 5034

Qy    661 caaaaaccaatcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5035 CAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAAGGAATCTCG 5094

Qy    721 caattaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtat 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5095 CAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTAT 5154

Qy    781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5155 TTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTA 5214

Qy    841 gaatcttccacaaagagcctttcagaaatttggggatcagcgttaccgaattttcacaagt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5215 GAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGT 5274

Qy    901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaattttatg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5275 TGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATG 5334

Qy    961 gaaaatatcatacaacccccctatccttgatgataaagagaaagcggagtttttgaaatct 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5335 GAAAATATCATACAACCCCCTATCCTTGATGATAAAGAGAAAGCGGAGTTTTTGAAATCT 5394

Qy   1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   5395 GCCAAACAATCTTTTGCAGGAATCATTATAGGGAATCAAATCCGAACGGATCAAAAGTTC 5454

Qy   1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaaatggagagcct 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5455 ATGGGCGTGTTTGATGAGTCCTTGAAAGAAAGGCAAGAAGCAGAAAAAAATGGAGAGCCT 5514

Qy   1141 actggtggggattggttggatatttttctctcatttatatttgacaaaaaacaatcttct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5515 ACTGGTGGGGATTGGTTGGATATTTTTCTCTCATTTATATTTGACAAAAAACAATCTTCT 5574

Qy   1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5575 GATGTCAAAGAAGCAATCAATCAAGAACCAGTTCCCCATGTCCAACCAGATATAGCCACT 5634

Qy   1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5635 ACCACCACCGACATACAAGGCTTACCGCCTGAAGCTAGAGATTTACTTGATGAAAGGGGT 5694

Qy   1321 aatttttctaaattcactcttggtgatatggaaatgttagatgttgagggagtcgctgac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5695 AATTTTTCTAAATTCACTCTTGGCGATATGGAAATGTTAGATGTTGAGGGAGTCGCTGAC 5754

Qy   1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtctttctgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5755 ATTGATCCCAATTACAAGTTCAATCAATTATTGATTCACAATAACGCTCTGTCTTCTGTG 5814

Qy   1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5815 TTAATGGGGAGTCATAATGGCATAGAACCTGAAAAAGTTTCATTGTTGTATGGGGGCAAT 5874

Qy   1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5875 GGTGGTCCTGGAGCTAGGCATGATTGGAACGCCACCGTTGGTTATAAAGACCAACAAGGC 5934

Qy   1561 aacaatgtggctacaataattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5935 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 5994

Qy   1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5995 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 6054

Qy   1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaattt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6055 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 6114

Qy   1741 cttgcacaaaaataatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6115 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCGAGAAAGAGAAGGAAAAATTCCGA 6174

Qy   1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
```



```
Db   6175 ACTGAGATTAAAGATTTCCAAAAAGACTCTAAGGCTTATTTAGACGCCCTAGGGAATGAT 6234

Qy   1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6235 CGTATTGCTTTTGTTTCTAAAAAAGACACAAAACATTCAGCTTTAATTACTGAGTTTGGT 6294

Qy   1921 aatgggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6295 AATGGGGATTTGAGCTACACTCTCAAAGATTATGGGAAAAAAGCAGATAAAGCTTTAGAT 6354

Qy   1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6355 AGGGAGAAAAATGTTACTCTTCAAGGTAGCCTAAAACATGATGGCGTGATGTTTGTTGAT 6414

Qy   2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6415 TATTCTAATTTCAAATACACCAACGCCTCCAAGAATCCCAATAAGGGTGTAGGCGTTACG 6474

Qy   2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6475 AATGGCGTTTCCCATTTAGAAGTAGGCTTTAACAAGGTAGCTATCTTTAATTTGCCTGAT 6534

Qy   2161 ttaaataatctcgctatcactagtttcgtaaggcggaatttagaggataaactaaccact 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6535 TTAAATAATCTCGCTATCACTAGTTTCGTAAGGCGGAATTTAGAGGATAAACTAACCACT 6594

Qy   2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6595 AAAGGATTGTCCCCACAAGAAGCTAATAAGCTTATCAAAGATTTTTTGAGCAGCAACAAA 6654

Qy   2281 gaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6655 GAATTGGTTGGAAAAACTTTAAACTTCAATAAAGCTGTAGCTGACGCTAAAAACACAGGC 6714

Qy   2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6715 AATTATGATGAAGTGAAAAAAGCTCAGAAAGATCTTGAAAAATCTCTAAGGAAACGAGAG 6774

Qy   2401 catttagagaaagaagtagagaaaaaatggagagcaaaagcggcaacaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6775 CATTTAGAGAAAGAAGTAGAGAAAAAATGGAGAGCAAAAGCGGCAACAAAAATAAAATG 6834

Qy   2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6835 GAAGCAAAAGCTCAAGCTAACAGCCAAAAAGATGAGATTTTTGCGTTGATCAATAAAGAG 6894

Qy   2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaagggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6895 GCTAATAGAGACGCAAGAGCAATCGCTTACGCTCAGAATCTTAAAGGCATCAAAAGGGAA 6954
```



```
Qy   2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6955 TTGTCTGATAAACTTGAAAATGTCAACAAGAATTTGAAAGACTTTGATAAATCTTTTGAT 7014

Qy   2641 gaattcaaaaatggcaaaaatgaaggatttcagcaaggcagaagaaaacactaaaagccctt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7015 GAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGCAGAAGAAACACTAAAAGCCCTT 7074

Qy   2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaacctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7075 AAAGGTTCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAAAGTTGAAAACCTT 7134

Qy   2761 aatgcagctttgaatgaattcaaaaatggcaaaaatgaaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7135 AATGCAGCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGTAACGCAA 7194

Qy   2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7195 GCAAAAAGCGACCTTGAAAATTCCGTTAAAGATGTGATCATCAATCAAAAGGTAACGGAT 7254

Qy   2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7255 AAAGTTGATAATCTCAATCAAGCGGTATCAGTGGCTAAAGCAACGGGTGATTTCAGTAGG 7314

Qy   2941 gtagaggcaagcgttagccgatctcaaaaatttctcaaaggagcaattggccgcaacaagct 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7315 GTAGAGCAAGCGTTAGCCGATCTCAAAAATTTCTCAAAGGAGCAATTGGCCGCAACAAGCT 7374

Qy   3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7375 CAAAAAAATGAAAGTCTCAATGCTAGAAAAAAATCTGAAATATATCAATCCGTTAAGAAT 7434

Qy   3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7435 GGTGTGAATGGAACCCTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACTCTTTCT 7494

Qy   3121 aaaaactttccggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7495 AAAAACTTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAAT 7554

Qy   3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7555 AACAATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGG 7614

Qy   3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7615 CAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAA 7674

Qy   3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7675 ATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC 7734
```

```
Qy   3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7735 CCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAA 7794

Qy   3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggtttt 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7795 GAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTT 7854

Qy   3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatccccatg 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7855 TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATG 7914

Qy   3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7915 AATCTATGGGTTGAAAGTGCAAAAAAAGTACCTGCTAGTTTGTCAGCGAAACTAGACAAT 7974

Qy   3601 tacgctactaacagcccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7975 TACGCTACTAACAGCCACATACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAA 8034

Qy   3661 aaagcgaccggcatgctaacgcaaaaaaacccctgagtggctcaagctcgtgaatgataag 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8035 AAAGCGACCGGCATGCTAACGCAAAAAAACCCTGAGTGGCTCAAGCTCGTGAATGATAAG 8094

Qy   3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaaattggcttcaac 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8095 ATAGTTGCGCATAATGTAGGAAGCGTTCCTTTGTCAGAGTATGATAAAATTGGCTTCAAC 8154

Qy   3781 cagaagaatatgaaagattattctgattcgttcaagttttccaccaagttgaacaatgct 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8155 CAGAAGAATATGAAAGATTATTCTGATTCGTTCAAGTTTTCCACCAAGTTGAACAATGCT 8214

Qy   3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcatttttctacagcatct 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8215 GTAAAAGACACTAATTCTGGCTTTACGCAATTTTTAACCAATGCATTTTCTACAGCATCT 8274

Qy   3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8275 TATTACTGCTTGGCGAGAGAAAATGCGGAGCATGGAATCAAGAACGTTAATACAAAAGGT 8334

Qy   3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccaccccttgcta 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8335 GGTTTCCAAAAATCTTAAAGGATTAAGGAATACCAAAAACGCAAAAACCACCCCTTGCTA 8394

Qy   4021 aaagcgagggggttttttaatactccttagcagaaatcccaatcgtctttagtatttggga 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8395 AAAGCGAGGGGTTTTTTAATACTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGA 8454

Qy   4081 tgaatgctaccaattcatggtatcatatcccccatacattcgtatctagcgtaggaagtgt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db   8455 TGAATGCTACCAATTCATGGTATCATATCCCCATACATTCGTATCTAGCGTAGGAAGTGT 8514

Qy   4141 gcaaagttacgccttttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8515 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 8574

Qy   4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctccctacgggca 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8575 AACTCTGTAAAATCCCTATTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 8634

Qy   4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8635 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 8694

Qy   4321 gacagacactaacgaaaggctttgttctttaaagtctgcatggatatttcctaccccaaa 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8695 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAA 8754

Qy   4381 aagacttaacccttttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8755 AAGACTTAACCCTTTGCTTAAAATTAAGTTTGATTGTGCTAGTGGGTTCGTGCTATAGTG 8814

Qy   4441 cgaaaattaattaagggttataaagagagcataaactagaaaaaaacaagtagctataaca 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8815 CGAAAATTAATTAAGGGTTATAAAGAGAGCATAAACTAGAAAAAACAAGTAGCTATAACA 8874

Qy   4501 aagatcaagttcaaaaaatcatagagctttttagagcaaattgatcgcgctcttaaccaaa 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8875 AAGATCAAGTTCAAAAAATCATAGAGCTTTTAGAGCAAATTGATCGCGCTCTTAACCAAA 8934

Qy   4561 gaaaaatcagaaaaaccatagaattatcacacacttataatgcccaaaaaagacgcttgc 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8935 GAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAAGACGCTTGC 8994

Qy   4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8995 GATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACTGTGG 9054

Qy   4681 atgcctttcaaggtgaagaggcagatattattatttattccaccgtgaaaacttgtggta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9055 ATGCCTTTCAAGGTGAAGAGGCAGATATTATTATTTATTCCACCGTGAAAACTTGTGGTA 9114

Qy   4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9115 ATCTTTCTTTCTTGCTAGATTCTAAACGCTTGAATGTGGCTATTTCTAGGGCAAAAGAAA 9174

Qy   4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9175 ATCTCATTTTTGTGGGTAAAAAGTCTTTCTTTGAGAATTTACGAAGCGATGAGAAGAATA 9234

Qy   4861 tctttagcgctattttgcaagtctgtagataggtaatcttttccaaagataatcattaga 4920
```



```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9235 TCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 9294

Qy   4921 cattcttcgcttcaaaacgctttcataaatctctctaaagcgctttataatcaacacaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9295 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACACAAT 9354

Qy   4981 accctatagtgtgagctatagccccttttgggaattgagttatttttgactttaaatttt 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9355 ACCCTTATAGTGTGAGCTATAGCCCCTTTTTGGGAATTGAGTTATTTTGACTTTAAATTT 9414

Qy   5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9415 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCC 9474

Qy   5101 gctcgcatgaaatttccactttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9475 GCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATTTTTTTTTAAGGGCGTATTTTTG 9534

Qy   5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggggtgttgaaagggc 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9535 CTGCAAATATCCTACAATAGCATCGCCCGAATGGATGAGTAGGGGGGGGTGTTGAAAGGGC 9594

Qy   5221 aaaatgctccataaaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9595 AAAATGCTCCATAAAATAGCCCTCAATTTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 9654

Qy   5281 taaaataatcacttcgggaaaatctttaagggagtgaaataataacgcatgcaagtttct 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9655 TAAAATAATCACTTCGGGAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTTTCT 9714

Qy   5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9715 AACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAAGAGAAGTGGCTAAA 9774

Qy   5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9775 TGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGGATC 9834

Qy   5461 gtcgcttttgtccctagcactaaaatagggggcgtttttatctttacttgtcgcttgatc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9835 GTCGCTTTTGTCCCTAGCACTAAAATAGGGGCGTTTTTATCTTTACTTGTCGCTTGATC 9894

Qy   5521 gctaaaatgcttggctcaatcacgcccacaatagggattttggaatgcttttgcatctct 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9895 GCTAAAATGCTTGGCTCAATCACGCCCACAATAGGGATTTTGGAATGCTTTTGCATCTCT 9954

Qy   5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9955 TCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGGT 10014
```



```
Qy   5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10015 TTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTGCCA 10074

Qy   5701 taaggcactctagccgtatcgccatagatgatttcatcaaataattgcgcttttaaa 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10075 TAAGGCACTCTAGCCGTATCGCCATAATAGATGATTTCATCAAATAATTGCGCTTTTAAA 10134

Qy   5761 aggctttttaaaaacgctaaaccctcccacaccgctatcaaaaaacgcctattttcatgaca 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10135 AGGCTTTTTAAAACGCTAAACCCTCCCACACCGCTATCAAAAACGCCTATTTTCATGACA 10194

Qy   5821 ctttttaatttaatgggattaattagggattttattttttcattcattaagtttaaaaat 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10195 CTTTTTTAATTTAATGGGATTAATTAGGGATTTTATTTTTCATTCATTAAGTTTAAAAAT 10254

Qy   5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
          |||||||||||||||||||||||||||||||||||||||||||||
Db  10255 TCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATAGACAAAGCTT 10299


RESULT   6
US-08-477-451-5/c
; Sequence 5, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
```

; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-477-451-5

Query Match 97.4%; Score 5772; DB 2; Length 10299;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5922; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 ctccatttaagcaactccatagaccactaaagaaactttttttgaggctatctttgaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5925 CTCCATTTTAAGCAACTCCATAGACCACTAAAGAAACTTTTTTTGAGGCTATCTTTGAAA 5866

Qy      61 atctgtcctattgatttgttttccatttgtttcccatgtggatcttgtggatcacaaac 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5865 ATCTGTCCTATTGATTTGTTTTCCATTTTGTTTCCCATGTGGATCTTGTGGATCACAAAC 5806

Qy     121 gcttaattatacatgctatagtaagcatgacacacaaaccaaactatttttagaacgctt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5805 GCTTAATTATACATGCTATAGTAAGCATGACACACAAACCAAACTATTTTTAGAACGCTT 5746

Qy     181 catgtgctcaccttgactaaccatttctccaaccatactttagcgttgcatttgatttct 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5745 CATGTGCTCACCTTGACTAACCATTTCTCCAACCATACTTTAGCGTTGCATTTGATTTCT 5686

Qy     241 tcaaaaagattcatttcttatttcttgttcttattaaagttctttcatttttagcaaattt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5685 TCAAAAAGATTCATTTCTTATTTCTTGTTCTTATTAAAGTTCTTTCATTTTAGCAAATTT 5626

Qy     301 ttgttaattgtgggtaaaaatgtgaatcgtcctagccttttagacgcctgcaacgatcggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5625 TTGTTAATTGTGGGTAAAAATGTGAATCGTCCTAGCCTTTAGACGCCTGCAACGATCGGG 5566

Qy     361 ctttttcaatattaataatgattaatgaaaaaaaaaaaaaaatgcttgatattgttgtat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5565 CTTTTTTCAATATTAATAATGATTAATGAAAAAAAAAAAAAATGCTTGATATTGTTGTAT 5506

Qy     421 aatgagaatgttcaaagacatgaattgactactcaagcgtgtagcgatttttagcagtct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5505 AATGAGAATGTTCAAAGACATGAATTGACTACTCAAGCGTGTAGCGATTTTTAGCAGTCT 5446

Qy     481 ttgacactaacaagatacgatagtatgaaactaggtatagtaaggagaaacaatgact 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5445 TTGACACTAACAAGATACCGATAGGTATGAAACTAGGTATAGTAAGGAGAAACAATGACT 5386
```



```
Qy     541 aacgaaacgcattgaccaacaaccacaaaccgaagcggcgtttttaacccgcagcaatttatc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5385 AACGAAACCATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATC 5326

Qy     601 aataatcttcaagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5325 AATAATCTTCAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGAT 5266

Qy     661 caaaaaccaatcgttgataagaacgataggagataacaggcaagcttttgaaggaatctcg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5265 CAAAAACCAATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAAGGAATCTCG 5206

Qy     721 caatttaagggaagaatactccaataaagcgatcaaaaatcctaccaaaaaagaatcagtat 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5205 CAATTAAGGGAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTAT 5146

Qy     781 ttttcagactttatcaataagagcaatgatttaatcaacaaagacaatctcattgatgta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5145 TTTTCAGACTTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTGATGTA 5086

Qy     841 gaatcttccacaaagagctttcagaaatttggggatcagcgttaccgaattttcacaagt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5085 GAATCTTCCACAAAGAGCTTTCAGAAATTTGGGGATCAGCGTTACCGAATTTTCACAAGT 5026

Qy     901 tgggtgtcccatcaaaacgatccgtctaaaatcaacacccgatcgatccgaaatttttatg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5025 TGGGTGTCCCATCAAAACGATCCGTCTAAAATCAACACCCGATCGATCCGAAATTTTATG 4966

Qy     961 gaaaatatcatacaaccccctatccttgatgataaagagaaagcggagatttttgaaatct 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4965 GAAAATATCATACAACCCCCTATCCTTGATGATAAAGAGAAAGCGGAGTTTTTGAAATCT 4906

Qy    1021 gccaaacaatcttttgcaggaatcattatagggaatcaaatccgaacggatcaaaagttc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4905 GCCAAACAATCTTTTGCAGGAATCATTATAGGGAATCAAATCCGAACGGATCAAAAGTTC 4846

Qy    1081 atgggcgtgtttgatgagtccttgaaagaaaggcaagaagcagaaaaaaatggagagagcct 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4845 ATGGGCGTGTTTGATGAGTCCTTGAAAGAAAGGCAAGAAGCAGAAAAAAATGGAGAGCCT 4786

Qy    1141 actggtgggggattggttggatatttttctctcatttatatttgacaaaaaaacaatcttct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4785 ACTGGTGGGGATTGGTTGGATATTTTTCTCTCATTTATATTTGACAAAAAACAATCTTCT 4726

Qy    1201 gatgtcaaagaagcaatcaatcaagaaccagttccccatgtccaaccagatatagccact 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4725 GATGTCAAAGAAGCAATCAATCAAGAACCAGTTCCCCATGTCCAACCAGATATAGCCACT 4666

Qy    1261 accaccaccgacatacaaggcttaccgcctgaagctagagatttacttgatgaaaggggt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4665 ACCACCACCGACATACAAGGCTTACCGCCTGAAGCTAGAGATTTACTTGATGAAAGGGGT 4606
```



```
Qy    1321 aatttttctaaattcactcttggcgatatggaaatgttagatgttgagggagtcgctgac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4605 AATTTTTCTAAATTCACTCTTGGCGATATGGAAATGTTAGATGTTGAGGGAGTCGCTGAC 4546

Qy    1381 attgatcccaattacaagttcaatcaattattgattcacaataacgctctgtcttctgtg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4545 ATTGATCCCAATTACAAGTTCAATCAATTATTGATTCACAATAACGCTCTGTCTTCTGTG 4486

Qy    1441 ttaatggggagtcataatggcatagaacctgaaaaagtttcattgttgtatgggggcaat 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4485 TTAATGGGGAGTCATAATGGCATAGAACCTGAAAAAGTTTCATTGTTGTATGGGGGCAAT 4426

Qy    1501 ggtggtcctggagctaggcatgattggaacgccaccgttggttataaagaccaacaaggc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4425 GGTGGTCCTGGAGCTAGGCATGATTGGAACGCCACCGTTGGTTATAAAGACCAACAAGGC 4366

Qy    1561 aacaatgtggctacaatgtaattaatgtgcatatgaaaaacggcagtggcttagtcatagca 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4365 AACAATGTGGCTACAATAATTAATGTGCATATGAAAAACGGCAGTGGCTTAGTCATAGCA 4306

Qy    1621 ggtggtgagaaagggattaacaaccctagtttttatctctacaaagaagaccaactcaca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4305 GGTGGTGAGAAAGGGATTAACAACCCTAGTTTTTATCTCTACAAAGAAGACCAACTCACA 4246

Qy    1681 ggctcacaacgagcattaagtcaagaagagatccaaaacaaaatagatttcatggaatttt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4245 GGCTCACAACGAGCATTAAGTCAAGAAGAGATCCAAAACAAAATAGATTTCATGGAATTT 4186

Qy    1741 cttgcacaaaatatgctaaattagacaacttgagcgagaaagagaaggaaaaattccga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4185 CTTGCACAAAATAATGCTAAATTAGACAACTTGAGCGAGAAAGAGAAGGAAAAATTCCGA 4126

Qy    1801 actgagattaaagatttccaaaaagactctaaggcttatttagacgccctagggaatgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4125 ACTGAGATTAAAGATTTCCAAAAAGACTCTAAGGCTTATTTAGACGCCCTAGGGAATGAT 4066

Qy    1861 cgtattgcttttgtttctaaaaaagacacaaaacattcagctttaattactgagtttggt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4065 CGTATTGCTTTTGTTTCTAAAAAAGACACAAAACATTCAGCTTTAATTACTGAGTTTGGT 4006

Qy    1921 aatggggatttgagctacactctcaaagattatgggaaaaaagcagataaagctttagat 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4005 AATGGGGATTTGAGCTACACTCTCAAAGATTATGGGAAAAAAGCAGATAAAGCTTTAGAT 3946

Qy    1981 agggagaaaaatgttactcttcaaggtagcctaaaacatgatggcgtgatgtttgttgat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3945 AGGGAGAAAAATGTTACTCTTCAAGGTAGCCTAAAACATGATGGCGTGATGTTTGTTGAT 3886

Qy    2041 tattctaatttcaaatacaccaacgcctccaagaatcccaataagggtgtaggcgttacg 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Db    3885 TATTCTAATTTCAAATACACCAACGCCTCCAAGAATCCCAATAAGGGTGTAGGCGTTACG 3826

Qy    2101 aatggcgtttcccatttagaagtaggctttaacaaggtagctatctttaatttgcctgat 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3825 AATGGCGTTTCCCATTTAGAAGTAGGCTTTAACAAGGTAGCTATCTTTAATTTGCCTGAT 3766

Qy    2161 ttaaatcatctcgctatcactagtttcgtaaggcggaatttagaggatcaaactaaccaat 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3765 TTAAATAATCTCGCTATCACTAGTTTCGTAAGGCGGAATTTAGAGGATCAAACTAACCAAT 3706

Qy    2221 aaaggattgtccccacaagaagctaataagcttatcaaagatttttgagcagcaacaaa 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3705 AAAGGATTGTCCCCACAAGAAGCTAATAAGCTTATCAAAGATTTTTTGAGCAGCAACAAA 3646

Qy    2281 gaattggttggaaaaactttaaacttcaataaagctgtagctgacgctaaaaacacaggc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3645 GAATTGGTTGGAAAAACTTTAAACTTCAATAAAGCTGTAGCTGACGCTAAAAACACAGGC 3586

Qy    2341 aattatgatgaagtgaaaaaagctcagaaagatcttgaaaaatctctaaggaaacgagag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3585 AATTATGATGAAGTGAAAAAAGCTCAGAAAGATCTTGAAAAATCTCTAAGGAAACGAGAG 3526

Qy    2401 catttagagaaagaagtagagaaaaaattggagagcaaaagcggcaacaaaaaataaaatg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3525 CATTTAGAGAAAGAAGTAGAGAAAAAATTGGAGAGCAAAAGCGGCAACAAAAATAAAATG 3466

Qy    2461 gaagcaaaagctcaagctaacagccaaaaagatgagatttttgcgttgatcaataaagag 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3465 GAAGCAAAAGCTCAAGCTAACAGCCAAAAAGATGAGATTTTTGCGTTGATCAATAAAGAG 3406

Qy    2521 gctaatagagacgcaagagcaatcgcttacgctcagaatcttaaaggcatcaaaaggggaa 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3405 GCTAATAGAGACGCAAGAGCAATCGCTTACGCTCAGAATCTTAAAGGCATCAAAAGGGAA 3346

Qy    2581 ttgtctgataaacttgaaaatgtcaacaagaatttgaaagactttgataaatcttttgat 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3345 TTGTCTGATAAACTTGAAAATGTCAACAAGAATTTGAAAGACTTTGATAAATCTTTTGAT 3286

Qy    2641 gaattcaaaaatggcaaaaataaggatttcagcaaggcagaagaaacactaaaagccctt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3285 GAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGCAGAAGAAACACTAAAAGCCCTT 3226

Qy    2701 aaaggttcggtgaaagatttaggtatcaatccagaatggatttcaaaagttgaaaacctt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3225 AAAGGTTCGGTGAAAGATTTAGGTATCAATCCAGAATGGATTTCAAAAGTTGAAAACCTT 3166

Qy    2761 aatgcagctttgaatgaattcaaaaatggcaaaaataaggatttcagcaaggtaacgcaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3165 AATGCAGCTTTGAATGAATTCAAAAATGGCAAAAATAAGGATTTCAGCAAGGTAACGCAA 3106

Qy    2821 gcaaaaagcgaccttgaaaattccgttaaagatgtgatcatcaatcaaaaggtaacggat 2880
```

```
Db  3105 GCAAAAAGCGACCTTGAAAATTCCGTTAAAGATGTGATCATCAATCAAAAGGTAACGGAT 3046
Qy  2881 aaagttgataatctcaatcaagcggtatcagtggctaaagcaacgggtgatttcagtagg 2940
Db  3045 AAAGTTGATAATCTCAATCAAGCGGTATCAGTGGCTAAAGCAACGGGTGATTTCAGTAGG 2986
Qy  2941 gtagagcaagcgttagccgatctcaaaaatttctcaaaggagcaattggcccaacaagct 3000
Db  2985 GTAGAGCAAGCGTTAGCCGATCTCAAAAATTTCTCAAAGGAGCAATTGGCCCAACAAGCT 2926
Qy  3001 caaaaaaatgaaagtctcaatgctagaaaaaaatctgaaatatatcaatccgttaagaat 3060
Db  2925 CAAAAAAATGAAAGTCTCAATGCTAGAAAAAAATCTGAAATATATCAATCCGTTAAGAAT 2866
Qy  3061 ggtgtgaatggaaccctagtcggtaatgggttatctcaagcagaagccacaactctttct 3120
Db  2865 GGTGTGAATGGAACCCTAGTCGGTAATGGGTTATCTCAAGCAGAAGCCACAACTCTTTCT 2806
Qy  3121 aaaaactttcggacatcaagaaagagttgaatgcaaaacttggaaatttcaataacaat 3180
Db  2805 AAAAACTTTTCGGACATCAAGAAAGAGTTGAATGCAAAACTTGGAAATTTCAATAACAAT 2746
Qy  3181 aacaataatggactcaaaaacgaacccatttatgctaaagttaataaaaagaaagcaggg 3240
Db  2745 AACAATAATGGACTCAAAAACGAACCCATTTATGCTAAAGTTAATAAAAAGAAAGCAGGG 2686
Qy  3241 caagcagctagccttgaagaacccatttacgctcaagttgctaaaaaggtaaatgcaaaa 3300
Db  2685 CAAGCAGCTAGCCTTGAAGAACCCATTTACGCTCAAGTTGCTAAAAAGGTAAATGCAAAA 2626
Qy  3301 attgaccgactcaatcaaatagcaagtggtttgggtgttgtagggcaagcagcgggcttc 3360
Db  2625 ATTGACCGACTCAATCAAATAGCAAGTGGTTTGGGTGTTGTAGGGCAAGCAGCGGGCTTC 2566
Qy  3361 cctttgaaaaggcatgataaagttgatgatctcagtaaggtagggctttcaaggaatcaa 3420
Db  2565 CCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAA 2506
Qy  3421 gaattggctcagaaaattgacaatctcaatcaagcggtatcagaagctaaagcaggttttt 3480
Db  2505 GAATTGGCTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTT 2446
Qy  3481 tttggcaatctagagcaaacgatagacaagctcaaagattctacaaaacacaatcccatg 3540
Db  2445 TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATG 2386
Qy  3541 aatctatgggttgaaagtgcaaaaaaagtacctgctagtttgtcagcgaaactagacaat 3600
Db  2385 AATCTATGGGTTGAAAGTGCAAAAAAGTACCTGCTAGTTTGTCAGCGAAACTAGACAAT 2326
```



```
Qy  3601 tacgctactaacagccacatacgcattaatagcaatatcaaaaatggagcaatcaatgaa 3660
Db  2325 TACGCTACTAACAGCCACATACGCATTAATAGCAATATCAAAAATGGAGCAATCAATGAA 2266
Qy  3661 aaagcgaccggcatgctaacgcaaaaaaaccctgagtggctcaagctcgtgaatgataag 3720
Db  2265 AAAGCGACCGGCATGCTAACGCAAAAAAACCCTGAGTGGCTCAAGCTCGTGAATGATAAG 2206
Qy  3721 atagttgcgcataatgtaggaagcgttcctttgtcagagtatgataaaattggcttcaac 3780
Db  2205 ATAGTTGCGCATAATGTAGGAAGCGTTCCTTTGTCAGAGTATGATAAAATTGGCTTCAAC 2146
Qy  3781 cagaagaatatgaaagattatttctgattcgttcaagttttccaccaagttgaacaatgct 3840
Db  2145 CAGAAGAATATGAAAGATTATTCTGATTCGTTCAAGTTTTCCACCAAGTTGAACAATGCT 2086
Qy  3841 gtaaaagacactaattctggctttacgcaatttttaaccaatgcattttctacagcatct 3900
Db  2085 GTAAAAGACACTAATTCTGGCTTTACGCAATTTTTAACCAATGCATTTTCTACAGCATCT 2026
Qy  3901 tattactgcttggcgagagaaaatgcggagcatggaatcaagaacgttaatacaaaaggt 3960
Db  2025 TATTACTGCTTGGCGAGAGAAAATGCGGAGCATGGAATCAAGAACGTTAATACAAAAGGT 1966
Qy  3961 ggtttccaaaaatcttaaaggattaaggaatacccaaaaacgcaaaaaccacccccttgcta 4020
Db  1965 GGTTTCCAAAAATCTTAAAGGATTAAGGAATACCAAAAACGCAAAAACCACCCCTTGCTA 1906
Qy  4021 aaagcgaggggttttttaatactccttagcagaaatcccaatggtctttagtatttgggga 4080
Db  1905 AAAGCGAGGGGTTTTTTAATACTCCTTAGCAGAAATCCCAATCGTCTTTAGTATTTGGGA 1846
Qy  4081 tgaatgctaccaattcatggtatcatatccccatacattcgtatctagcgtaggaagtgt 4140
Db  1845 TGAATGCTACCAATTCATGGTATCATATCCCCATACATTCGTATCTAGCGTAGGAAGTGT 1786
Qy  4141 gcaaagttacgccttggagatatgatgtgtgagacctgtagggaatgcgttggagctca 4200
Db  1785 GCAAAGTTACGCCTTTGGAGATATGATGTGTGAGACCTGTAGGGAATGCGTTGGAGCTCA 1726
Qy  4201 aactctgtaaaatccctattatagggacacagagtgagaaccaaactctcccctacgggca 4260
Db  1725 AACTCTGTAAAATCCCTATTATAGGGACACAGAGTGAGAACCAAACTCTCCCTACGGGCA 1666
Qy  4261 acatcagcctaggaagcccaatcgtctttagcggttgggcacttcaccttaaaatatccc 4320
Db  1665 ACATCAGCCTAGGAAGCCCAATCGTCTTTAGCGGTTGGGCACTTCACCTTAAAATATCCC 1606
Qy  4321 gacagacactaacgaaaggcttttgttctttaaagtctgcatggatatttcctaccccaaa 4380
Db  1605 GACAGACACTAACGAAAGGCTTTGTTCTTTAAAGTCTGCATGGATATTTCCTACCCCAAA 1546
```



```
Qy  4381 aagacttaacccctttgcttaaaattaagtttgattgtgctagtgggttcgtgctatagtg 4440
Db  1545 AAGACTTAACCCTTTGCTTAAAATTAAGTTTGATTGTGCTAGTGGGTTCGTGCTATAGTG 1486
Qy  4441 cgaaaattaattaagggttatataagagagcataaactagaaaaaacaagtagctataaca 4500
Db  1485 CGAAAATTAATTAAGGGTTATAAAGAGAGCATAAACTAGAAAAAACAAGTAGCTATAACA 1426
Qy  4501 aagatcaagttcaaaaaatcatagagcttttagagcaaattgatcgcgctcttaaccaaa 4560
Db  1425 AAGATCAAGTTCAAAAAATCATAGAGCTTTTAGAGCAAATTGATCGCGCTCTTAACCAAA 1366
Qy  4561 gaaaaatcagaaaaaccataggaattatcacaccttataatgcccaaaaagacgcttgc 4620
Db  1365 GAAAAATCAGAAAAACCATAGGAATTATCACACCTTATAATGCCCAAAAAGACGCTTGC 1306
Qy  4621 gatcagaagtggaaaaatacggcttcaagaattttgatgagctcaaaatagacactgtgg 4680
Db  1305 GATCAGAAGTGGAAAAATACGGCTTCAAGAATTTTGATGAGCTCAAAATAGACACTGTGG 1246
Qy  4681 atgcctttcaaggtgaagaggcagatattattatttattccacccgtgaaaacttgtggta 4740
Db  1245 ATGCCTTTCAAGGTGAAGAGGCAGATATTATTATTTATTCCACCGTGAAAACTTGTGGTA 1186
Qy  4741 atctttctttcttgctagattctaaacgcttgaatgtggctatttctagggcaaaagaaa 4800
Db  1185 ATCTTTCTTTCTTGCTAGATTCTAAACGCTTGAATGTGGCTATTTCTAGGGCAAAAGAAA 1126
Qy  4801 atctcatttttgtgggtaaaaagtctttctttgagaatttatgaagcgatgagaagaata 4860
Db  1125 ATCTCATTTTTGTGGGTAAAAAGTCTTTCTTTGAGAATTTACGAAGCGATGAGAAGAATA 1066
Qy  4861 tctttagcgctattttgcaagtctgtagataggtaatcttttccaaagatgatcattagga 4920
Db  1065 TCTTTAGCGCTATTTTGCAAGTCTGTAGATAGGTAATCTTTTCCAAAGATAATCATTAGA 1006
Qy  4921 catttcttcgcttccaaaacgctttcataaatctctctcaaagcgctttataatcaacacaat 4980
Db  1005 CATTCTTCGCTTCAAAACGCTTTCATAAATCTCTCTAAAGCGCTTTATAATCAACACAAT 946
Qy  4981 accctatagtgtgagctatagccccttttttgggaattgagttatttttgacttaaattt 5040
Db  945 ACCCTTATAGTGTGAGCTATAGCCCCTTTTTGGGAATTGAGTTATTTTGACTTAAATTT 886
Qy  5041 ttattagcgttacaatttgagccattctttagcttgtttttctagccagatcacatcgcc 5100
Db  885 TTATTAGCGTTACAATTTGAGCCATTCTTTAGCTTGTTTTTCTAGCCAGATCACATCGCC 826
Qy  5101 gctcgcatgaaattccacctttagggaatgcgtgtgcatttttttttaagggcgtatttttg 5160
```



```
Db   825 GCTCGCATGAAATTCCACTTTAGGGAATGCGTGTGCATTTTTTTTAAGGGCGTATTTTTG 766
Qy  5161 ctgcaaatatcctacaatagcatcgcccgaatggatgagtaggggggtgttgaaagggc 5220
Db   765 CTGCAAATATCCTACAATAGCATCGCCCGAATGGATGAGTAGGGGGGGTGTTGAAAGGGC 706
Qy  5221 aaaatgctccataaaaatagccctcaatttttgagcgattaagggaaaatgcgtgcaacc 5280
Db   705 AAAATGCTCCATAAAATAGCCCTCAATTTTTTGAGCGATTAAGGGAAAATGCGTGCAACC 646
Qy  5281 taaaatgatcacttcgggaaaatctttaagggagtgaaatgataacgcatgcaagtttct 5340
Db   645 TAAAATAATCACTTCGGGAAAATCTTTAAGGGAGTGAAATAATAACGCATGCAAGTTTCT 586
Qy  5341 aacaattcgccctctaaaatactttcttcaatcaaaggcacaaaaagagaagtggctaaa 5400
Db   585 AACAATTCGCCCTCTAAAATACTTTCTTCAATCAAAGGCACAAAAAGAGAAGTGGCTAAA 526
Qy  5401 tgcgaaacattcaaatagccttgttgtttcagggcattgtcataagcgttggattggatc 5460
Db   525 TGCGAAACATTCAAATAGCCTTGTTGTTTCAGGGCATTGTCATAAGCGTTGGATTGGATC 466
Qy  5461 gtcgcttttgtccctagcactaaaatagggggcgtttttatctttttacttgtcgcttgatc 5520
Db   465 GTCGCTTTTGTCCCTAGCACTAAAATAGGGGGCGTTTTTATCTTTTACTTGTCGCTTGATC 406
Qy  5521 gctaaaatgcttggctcaatcacgcccacaatagggattttggaatgcttttgcatctct 5580
Db   405 GCTAAAATGCTTGGCTCAATCACGCCCACAATAGGGATTTTGGAATGCTTTTGCATCTCT 346
Qy  5581 tctaaagctagagcgctcgctgtgttgcatgccacaatcaataattcaatctggtgcggt 5640
Db   345 TCTAAAGCTAGAGCGCTCGCTGTGTTGCATGCCACAATCAATAATTCAATCTGGTGCGGT 286
Qy  5641 ttgaaaaaatccaaagcctctaagccaaattgcttgatcgtagtggggtctttagtgcca 5700
Db   285 TTGAAAAAATCCAAAGCCTCTAAGCCAAATTGCTTGATCGTAGTGGGGTCTTTAGTGCCA 226
Qy  5701 taaggcactctagccgtatcgccataatagatgatttcatcaaataattgcgcttttaaa 5760
Db   225 TAAGGCACTCTAGCGCTATCGCCATAATAGATGATTTCATCAAATAATTGCGCTTTTAAA 166
Qy  5761 aggctttttaaaacgctaaacccctcccacaccgctatcaaaaacgcctatttttcatgaca 5820
Db   165 AGGCTTTTTAAAACGCTAAACCCTCCCACACCGCTATCAAAAACGCCTATTTTCATGACA 106
Qy  5821 ctttttttaatttaatgggattaattagggatttattttttcattcattaagtttaaaaat 5880
Db   105 CTTTTTTAATTTAATGGGATTAATTAGGGATTTTATTTTTCATTCATTAAGTTTAAAAAT 46
Qy  5881 tcttcattgtccttagtttgttgcattttagaatagacaaagctt 5925
```

```
                     ||||||||||||||||||||||||||||||||||||||||||||
Db      45 TCTTCATTGTCCTTAGTTTGTTGCATTTTAGAATAGACAAAGCTT 1
```

```
RESULT   7
US-08-053-614-1
; Sequence 1, Application US/08053614
; Patent No. 5403924
; GENERAL INFORMATION:
;   APPLICANT:  COVER, TIMOTHY L.
;   APPLICANT:  BLASER, MARTIN J.
;   APPLICANT:  TUMMURU, MURALI K.R.
;   TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;   TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;     STREET:  133 Carnegie Way, Suite 400
;     CITY:  Atlanta
;     STATE:  Georgia
;     COUNTRY:  USA
;     ZIP:  30303
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/053,614
;     FILING DATE:  19930426
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  SPRATT, GWENDOLYN D.
;     REGISTRATION NUMBER:  36,016
;     REFERENCE/DOCKET NUMBER:  2200.009
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  404/688-0770
;     TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  3648 base pairs
;     TYPE:  NUCLEIC ACID
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   ORIGINAL SOURCE:
;     ORGANISM:  Helicobacter pylori
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  1072..3648
US-08-053-614-1
```



```
  Query Match             4.0%;  Score 235;  DB 1;  Length 3648;
  Best Local Similarity   99.7%;  Pred. No. 9e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372
```

```
RESULT   8
US-08-316-397B-1
; Sequence 1, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
;   APPLICANT:  COVER, TIMOTHY L.
;   APPLICANT:  BLASER, MARTIN J.
;   APPLICANT:  HARRY KLEANTHOUS
;   APPLICANT:  TUMMURU, MURALI K.R.
;   TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;   TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;     STREET:  127 Peachtree Street, Suite 1200
;     CITY:  Atlanta
;     STATE:  Georgia
;     COUNTRY:  USA
;     ZIP:  30303
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
```



```
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/316,397B
;     FILING DATE:
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  SPRATT, GWENDOLYN D.
;     REGISTRATION NUMBER:  36,016
;     REFERENCE/DOCKET NUMBER:  2200.030
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  404/688-0770
;     TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  3648 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   ORIGINAL SOURCE:
;     ORGANISM:  Helicobacter pylori
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  1072..3648
US-08-316-397B-1
```


```
  Query Match             4.0%;  Score 235;  DB 1;  Length 3648;
  Best Local Similarity   99.7%;  Pred. No. 9e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372
```



```
RESULT   9
US-09-034-306-1
; Sequence 1, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
;   APPLICANT:  COVER, TIMOTHY L.
;   APPLICANT:  BLASER, MARTIN J.
;   APPLICANT:  HARRY KLEANTHOUS
;   APPLICANT:  TUMMURU, MURALI K.R.
;   TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;   TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;     STREET:  127 Peachtree Street, Suite 1200
;     CITY:  Atlanta
;     STATE:  Georgia
;     COUNTRY:  USA
;     ZIP:  30303
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/034,306
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/316,397
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  SPRATT, GWENDOLYN D.
;     REGISTRATION NUMBER:  36,016
;     REFERENCE/DOCKET NUMBER:  2200.030
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  404/688-0770
;     TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  3648 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   ORIGINAL SOURCE:
;     ORGANISM:  Helicobacter pylori
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  1072..3648
US-09-034-306-1
```

```
  Query Match             4.0%;  Score 235;  DB 2;  Length 3648;
  Best Local Similarity   99.7%;  Pred. No. 9e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  10
US-09-259-437-1
; Sequence 1, Application US/09259437
; Patent No. 6153390
;  GENERAL INFORMATION:
;    APPLICANT:  COVER, TIMOTHY L.
;    APPLICANT:  BLASER, MARTIN J.
;    APPLICANT:  HARRY KLEANTHOUS
;    APPLICANT:  TUMMURU, MURALI K.R.
;    TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  127 Peachtree Street, Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```



```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/259,437
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/316,397
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SPRATT, GWENDOLYN D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  2200.030
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3648 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Helicobacter pylori
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1072..3648
US-09-259-437-1


  Query Match             4.0%;  Score 235;  DB 3;  Length 3648;
  Best Local Similarity   99.7%;  Pred. No. 9e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372
```



```
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  11
PCT-US93-09782-1
; Sequence 1, Application PC/TUS9309782
;  GENERAL INFORMATION:
;    APPLICANT:  COVER, TIMOTHY L.
;    APPLICANT:  BLASER, MARTIN J.
;    APPLICANT:  TUMMURU, MURALI K.R.
;    TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  133 Carnegie Way, Suite 400
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/09782
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SPRATT, GWENDOLYN D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  2200.009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3648 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Helicobacter pylori
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1072..3648
PCT-US93-09782-1
```



```
  Query Match             4.0%;  Score 235;  DB 4;  Length 3648;
  Best Local Similarity   99.7%;  Pred. No. 9e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  12
US-08-053-614-3
; Sequence 3, Application US/08053614
; Patent No. 5403924
;  GENERAL INFORMATION:
;    APPLICANT:  COVER, TIMOTHY L.
;    APPLICANT:  BLASER, MARTIN J.
;    APPLICANT:  TUMMURU, MURALI K.R.
;    TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  133 Carnegie Way, Suite 400
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/053,614
;      FILING DATE:  19930426
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SPRATT, GWENDOLYN D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  2200.009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4821 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1072..4614
US-08-053-614-3


  Query Match             4.0%;  Score 235;  DB 1;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 8.8e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggctttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  13
US-08-316-397B-3
; Sequence 3, Application US/08316397B
; Patent No. 5733740
;  GENERAL INFORMATION:
```



```
;    APPLICANT:  COVER, TIMOTHY L.
;    APPLICANT:  BLASER, MARTIN J.
;    APPLICANT:  HARRY KLEANTHOUS
;    APPLICANT:  TUMMURU, MURALI K.R.
;    TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  127 Peachtree Street, Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/316,397B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SPRATT, GWENDOLYN D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  2200.030
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1072..4614
US-08-316-397B-3


  Query Match             4.0%;  Score 235;  DB 1;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 8.8e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggctttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146
```



```
Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  14
US-09-034-306-3
; Sequence 3, Application US/09034306
; Patent No. 5876943
;  GENERAL INFORMATION:
;    APPLICANT:  COVER, TIMOTHY L.
;    APPLICANT:  BLASER, MARTIN J.
;    APPLICANT:  HARRY KLEANTHOUS
;    APPLICANT:  TUMMURU, MURALI K.R.
;    TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  127 Peachtree Street, Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/034,306
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/316,397
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SPRATT, GWENDOLYN D.
;      REGISTRATION NUMBER:  36,016
```



```
;      REFERENCE/DOCKET NUMBER:  2200.030
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1072..4614
US-09-034-306-3


  Query Match             4.0%;  Score 235;  DB 2;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 8.8e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggctttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaataaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATAAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  15
US-09-259-437-3
; Sequence 3, Application US/09259437
; Patent No. 6153390
;  GENERAL INFORMATION:
;    APPLICANT:  COVER, TIMOTHY L.
;    APPLICANT:  BLASER, MARTIN J.
;    APPLICANT:  HARRY KLEANTHOUS
;    APPLICANT:  TUMMURU, MURALI K.R.
```

```
;    TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  127 Peachtree Street, Suite 1200
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/259,437
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/316,397
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SPRATT, GWENDOLYN D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  2200.030
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1072..4614
US-09-259-437-3


  Query Match             4.0%;  Score 235;  DB 3;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 8.8e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaatgaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATGAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  16
PCT-US93-09782-3
; Sequence 3, Application PC/TUS9309782
;  GENERAL INFORMATION:
;    APPLICANT:  COVER, TIMOTHY L.
;    APPLICANT:  BLASER, MARTIN J.
;    APPLICANT:  TUMMURU, MURALI K.R.
;    TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  133 Carnegie Way, Suite 400
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/09782
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SPRATT, GWENDOLYN D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  2200.009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
```



```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaatgaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATGAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  16
PCT-US93-09782-3
; Sequence 3, Application PC/TUS9309782
;  GENERAL INFORMATION:
;    APPLICANT:  COVER, TIMOTHY L.
;    APPLICANT:  BLASER, MARTIN J.
;    APPLICANT:  TUMMURU, MURALI K.R.
;    TITLE OF INVENTION:  THE tagA GENE AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  PREDISPOSITION TO PEPTIC ULCERATION
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  133 Carnegie Way, Suite 400
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/09782
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SPRATT, GWENDOLYN D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  2200.009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
```



```
;      LENGTH:  4821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1072..4614
PCT-US93-09782-3


  Query Match             4.0%;  Score 235;  DB 4;  Length 4821;
  Best Local Similarity   99.7%;  Pred. No. 8.8e-89;
  Matches  285;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     550 attgaccaacaaccacaaaccgaagcggcttttaacccgcagcaatttatcaataatctt 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1087 ATTGACCAACAACCACAAACCGAAGCGGCTTTTAACCCGCAGCAATTTATCAATAATCTT 1146

Qy     610 caagtagcttttcttaaagttgataacgctgtcgcttcatacgatcctgatcaaaaacca 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1147 CAAGTAGCTTTTCTTAAAGTTGATAACGCTGTCGCTTCATACGATCCTGATCAAAAACCA 1206

Qy     670 atcgttgataagaacgatagggataacaggcaagcttttgaaggaatctcgcaattaagg 729
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    1207 ATCGTTGATAAGAACGATAGGGATAACAGGCAAGCTTTTGAGGGAATCTCGCAATTAAGG 1266

Qy     730 gaagaatactccaatgaagcgatcaaaaatcctaccaaaaagaatcagtatttttcagac 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1267 GAAGAATACTCCAATGAAGCGATCAAAAATCCTACCAAAAAGAATCAGTATTTTTCAGAC 1326

Qy     790 tttatcaataagagcaatgatttaatcaacaaagacaatctcattg 835
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1327 TTTATCAATAAGAGCAATGATTTAATCAACAAAGACAATCTCATTG 1372


RESULT  17
US-08-965-903B-5
; Sequence 5, Application US/08965903B
; Patent No. 6060275
;  GENERAL INFORMATION:
;    APPLICANT:  Hacohen, Nir
;    APPLICANT:  Krasnow, Mark A.
;    TITLE OF INVENTION:  SPROUTY PROTEIN AND CODING
;    TITLE OF INVENTION:  SEQUENCE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Ave., Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
```



```
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/965,903B
;      FILING DATE:  07-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/030232
;      FILING DATE:  07-NOV-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Petithory, Joanne R
;      REGISTRATION NUMBER:  42,995
;      REFERENCE/DOCKET NUMBER:  8600-0177.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-0880
;      TELEFAX:  650-324-0960
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1604 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  h-spryl cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...417
;      OTHER INFORMATION:
US-08-965-903B-5


  Query Match             0.4%;  Score 21;  DB 3;  Length 1604;
  Best Local Similarity  100.0%;  Pred. No. 8.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaaaaa 402
           |||||||||||||||||||||
Db    1580 ATTAATGAAAAAAAAAAAAAA 1600


RESULT  18
US-09-012-198-6/c
; Sequence 6, Application US/09012198
; Patent No. 6124271
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  P.L. Iversen, et al.
;    TITLE OF INVENTION:  Method and Conjugate for Treating
;    TITLE OF INVENTION:  H. Pylori Infection
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Ave.  Ste.250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/012,198
;      FILING DATE:  23-JAN-1998
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/848,844
;      FILING DATE:  30-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/036,366
;      FILING DATE:  24-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dehlinger, Peter J
;      REGISTRATION NUMBER:  28006
;      REFERENCE/DOCKET NUMBER:  0450-0019.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650 324-0880
;      TELEFAX:  650 324-0960
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-012-198-6


  Query Match             0.3%;  Score 20;  DB 3;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    524 aaggagaaacaatgactaac 543
          ||||||||||||||||||||
Db     20 AAGGAGAAACAATGACTAAC 1
```



```
RESULT  19
US-08-377-687-50
; Sequence 50, Application US/08377687
; Patent No. 5538525
;  GENERAL INFORMATION:
;    APPLICANT:  BROEKAERT, WILLEM F.
;    APPLICANT:  CAMMUE, BRUNO P.A.
;    APPLICANT:  OSBORN, RUPERT W.
;    APPLICANT:  REES, SARAH B.
;    APPLICANT:  TERRAS, FRANKY R.G.
;    APPLICANT:  VANDERLEYDEN, JOZEF
;    TITLE OF INVENTION:  BIOCIDAL PROTEINS
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/377,687
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/002,480
;      FILING DATE:  04-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  99042/SEE.36525/US/A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  284 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..108
US-08-377-687-50
```



```
  Query Match             0.3%;  Score 20;  DB 1;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5316 gaaataataacgcatgcaag 5335
          ||||||||||||||||||||
Db      2 GAAATAATAACGCATGCAAG 21


RESULT  20
US-08-777-192-50
; Sequence 50, Application US/08777192
; Patent No. 5824869
;  GENERAL INFORMATION:
;    APPLICANT:  BROEKAERT, WILLEM F.
;    APPLICANT:  CAMMUE, BRUNO P.A.
;    APPLICANT:  OSBORN, RUPERT W.
;    APPLICANT:  REES, SARAH B.
;    APPLICANT:  TERRAS, FRANKY R.G.
;    APPLICANT:  VANDERLEYDEN, JOZEF
;    TITLE OF INVENTION:  BIOCIDAL PROTEINS
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/777,192
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/002,480
;      FILING DATE:  04-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  99042/SEE.36525/US/A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;  INFORMATION FOR SEQ ID NO:  50:
```



```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  284 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..108
US-08-777-192-50


  Query Match             0.3%;  Score 20;  DB 1;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5316 gaaataataacgcatgcaag 5335
          ||||||||||||||||||||
Db      2 GAAATAATAACGCATGCAAG 21


RESULT  21
US-08-839-008-1
; Sequence 1, Application US/08839008
; Patent No. 5916758
;  GENERAL INFORMATION:
;    APPLICANT:  Hurle, Mark R
;    APPLICANT:  McDonnell, Peter C
;    APPLICANT:  McNulty, Dean E
;    APPLICANT:  Rosen, Craig A
;    APPLICANT:  Siemens, Ivo R
;    APPLICANT:  Young, Peter R
;    APPLICANT:  Yue, Tian-Li
;    TITLE OF INVENTION:  Smooth Muscle Cell-Derived Migration Factor
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/839,008
;      FILING DATE:  23-APR-1997
;      CLASSIFICATION:  435
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/563,697
;       FILING DATE:  28-NOV-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Baumeister, Kirk
;       REGISTRATION NUMBER:  33,833
;       REFERENCE/DOCKET NUMBER:  P50384
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610-270-5096
;       TELEFAX:  610-270-5090
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1537 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-839-008-1


  Query Match             0.3%;  Score 20;  DB 2;  Length 1537;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    1505 TTAATGAAAAAAAAAAAAAA 1524


RESULT  22
US-08-036-210-6
; Sequence 6, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
;     APPLICANT:  Moller, Niels P.H.
;     APPLICANT:  Moller, Karin B.
;     APPLICANT:  Ullrich, Axel
;     TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;     TITLE OF INVENTION:  PHOSPHATASE
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PENNIE & EDMONDS
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
```



```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/036,210
;       FILING DATE:  23-MAR-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  7683-025
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-790-9090
;       TELEFAX:  212-869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2173 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1630
;       OTHER INFORMATION:  /note= "N=x=unknown nucleotide"
US-08-036-210-6


  Query Match             0.3%;  Score 20;  DB 1;  Length 2173;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2138 TTAATGAAAAAAAAAAAAAA 2157


RESULT  23
US-08-449-609-6
; Sequence 6, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
;     APPLICANT:  Moller, Niels P.H.
;     APPLICANT:  Moller, Karin B.
;     APPLICANT:  Ullrich, Axel
;     TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;     TITLE OF INVENTION:  PHOSPHATASE
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PENNIE & EDMONDS
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
```



```
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/449,609
;       FILING DATE:  24-MAY-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/036,210
;       FILING DATE:  23-MAR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  7683-025
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-790-9090
;       TELEFAX:  212-869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2173 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1630
;       OTHER INFORMATION:  /note= "N=x=unknown nucleotide"
US-08-449-609-6


  Query Match             0.3%;  Score 20;  DB 2;  Length 2173;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2138 TTAATGAAAAAAAAAAAAAA 2157


RESULT  24
US-08-036-210-10
; Sequence 10, Application US/08036210
; Patent No. 5585233
;  GENERAL INFORMATION:
```



```
;     APPLICANT:  Moller, Niels P.H.
;     APPLICANT:  Moller, Karin B.
;     APPLICANT:  Ullrich, Axel
;     TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;     TITLE OF INVENTION:  PHOSPHATASE
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PENNIE & EDMONDS
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/036,210
;       FILING DATE:  23-MAR-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  7683-025
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-790-9090
;       TELEFAX:  212-869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2309 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1766
;       OTHER INFORMATION:  /note= "N=x=unknown nucleotide"
US-08-036-210-10


  Query Match             0.3%;  Score 20;  DB 1;  Length 2309;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2274 TTAATGAAAAAAAAAAAAAA 2293
```

```
RESULT  25
US-08-449-609-10
; Sequence 10, Application US/08449609
; Patent No. 5952212
;  GENERAL INFORMATION:
;    APPLICANT:  Moller, Niels P.H.
;    APPLICANT:  Moller, Karin B.
;    APPLICANT:  Ullrich, Axel
;    TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/449,609
;      FILING DATE:  24-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/036,210
;      FILING DATE:  23-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7683-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2309 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1766
;      OTHER INFORMATION:  /note= "N=x=unknown nucleotide"
```



```
US-08-449-609-10

  Query Match             0.3%;  Score 20;  DB 2;  Length 2309;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2274 TTAATGAAAAAAAAAAAAAA 2293


RESULT  26
PCT-US91-04541-1/c
; Sequence 1, Application PC/TUS9104541
;  GENERAL INFORMATION:
;    APPLICANT:  Oncogene Science Inc.
;    TITLE OF INVENTION:  TISSUE DERIVED TUMOR GROWTH INHIBITORS          METHODS OF PR
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/04541
;      FILING DATE:  19910625
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  1919/22669-F-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)977-9550
;      TELEFAX:  (212)664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2529 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  N
```



```
;    ANTI-SENSE:  N
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Transforming Growth Factor-Beta 3
;    IMMEDIATE SOURCE:
;      CLONE:  TGF-B3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  263..1498
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  263..1498
;      OTHER INFORMATION:
PCT-US91-04541-1


  Query Match             0.3%;  Score 20;  DB 4;  Length 2529;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db    1981 AAAAAAAAAAAAAATGCTTG 1962


RESULT  27
5262319-1/c
;Patent No. 5262319
;    APPLICANT: IWATA, KENNETH K.;FOULKES, J.GORDON;DIJKE,
;PETER T.;HALEY, JOHN D.
;    TITLE OF INVENTION: METHOD FOR OBBTAINING BONE MARROW FREE
;OF TUMOR CELLS USING TRANSFORMING GROWTH FACTOR B3
;    NUMBER OF SEQUENCES: 9
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/543,341
;      FILING DATE: 25-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 353,410
;      FILING DATE: 17-MAY-1989
;      APPLICATION NUMBER: 183,410
;      FILING DATE: 20-APR-1988
;      APPLICATION NUMBER: 111,022
;      FILING DATE: 20-OCT-1987
;      APPLICATION NUMBER: 922,121
;      FILING DATE: 20-OCT-1986
;      APPLICATION NUMBER: 847,931
;      FILING DATE: 07-APR-1986
;      APPLICATION NUMBER: 725,003
;      FILING DATE: 19-APR-1985
;SEQ ID NO:1:
```



```
;      LENGTH: 2529
5262319-1


  Query Match             0.3%;  Score 20;  DB 5;  Length 2529;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgcttg 408
           ||||||||||||||||||||
Db    1981 AAAAAAAAAAAAAATGCTTG 1962


RESULT  28
US-08-036-210-14
; Sequence 14, Application US/08036210
; Patent No. 5585233
;  GENERAL INFORMATION:
;    APPLICANT:  Moller, Niels P.H.
;    APPLICANT:  Moller, Karin B.
;    APPLICANT:  Ullrich, Axel
;    TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/036,210
;      FILING DATE:  23-MAR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7683-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2692 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  2149
;       OTHER INFORMATION:  /note= "N=x=unknown nucleotide"
US-08-036-210-14


  Query Match             0.3%;  Score 20;  DB 1;  Length 2692;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2657 TTAATGAAAAAAAAAAAAAA 2676


RESULT  29
US-08-449-609-14
; Sequence 14, Application US/08449609
; Patent No. 5952212
;  GENERAL INFORMATION:
;     APPLICANT:  Moller, Niels P.H.
;     APPLICANT:  Moller, Karin B.
;     APPLICANT:  Ullrich, Axel
;     TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;     TITLE OF INVENTION:  PHOSPHATASE
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PENNIE & EDMONDS
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/449,609
;       FILING DATE:  24-MAY-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/036,210
;       FILING DATE:  23-MAR-1993
;     ATTORNEY/AGENT INFORMATION:
```



```
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  7683-025
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-790-9090
;       TELEFAX:  212-869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2692 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  2149
;       OTHER INFORMATION:  /note= "N=x=unknown nucleotide"
US-08-449-609-14


  Query Match             0.3%;  Score 20;  DB 2;  Length 2692;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    2657 TTAATGAAAAAAAAAAAAAA 2676


RESULT  30
US-08-036-210-21
; Sequence 21, Application US/08036210
; Patent No. 5505233
;  GENERAL INFORMATION:
;     APPLICANT:  Moller, Niels P.H.
;     APPLICANT:  Moller, Karin B.
;     APPLICANT:  Ullrich, Axel
;     TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;     TITLE OF INVENTION:  PHOSPHATASE
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PENNIE & EDMONDS
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
```



```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/036,210
;       FILING DATE:  23-MAR-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  7683-025
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-790-9090
;       TELEFAX:  212-869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3973 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3430
;       OTHER INFORMATION:  /note= "N=x=unknown nucleotide"
US-08-036-210-21


  Query Match             0.3%;  Score 20;  DB 1;  Length 3973;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    3938 TTAATGAAAAAAAAAAAAAA 3957


RESULT  31
US-08-449-609-21
; Sequence 21, Application US/08449609
; Patent No. 5952212
;  GENERAL INFORMATION:
;     APPLICANT:  Moller, Niels P.H.
;     APPLICANT:  Moller, Karin B.
;     APPLICANT:  Ullrich, Axel
;     TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;     TITLE OF INVENTION:  PHOSPHATASE
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PENNIE & EDMONDS
;       STREET:  1155 Avenue of the Americas
```



```
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/449,609
;       FILING DATE:  24-MAY-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/036,210
;       FILING DATE:  23-MAR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  7683-025
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-790-9090
;       TELEFAX:  212-869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3973 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3430
;       OTHER INFORMATION:  /note= "N=x=unknown nucleotide"
US-08-449-609-21


  Query Match             0.3%;  Score 20;  DB 2;  Length 3973;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     383 ttaatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    3938 TTAATGAAAAAAAAAAAAAA 3957


RESULT  32
US-08-126-594-13/c
; Sequence 13, Application US/08126594
; Patent No. 5482845
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Soares, M. Bento
;    APPLICANT:  Efstratiadis, Algiris
;    TITLE OF INVENTION:  METHOD FOR CONSTRUCTION OF NORMALIZED
;    TITLE OF INVENTION:  cDNA LIBRARIES
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White, c/o Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/126,594
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  42840/JPW/AKC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-126-594-13


  Query Match             0.3%;  Score 19;  DB 1;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgctt 407
           ||||||||||||||||||
Db      32 AAAAAAAAAAAAAATGCTT 14
```



```
RESULT  33
US-08-465-811A-13/c
; Sequence 13, Application US/08465811A
; Patent No. 5637685
;  GENERAL INFORMATION:
;    APPLICANT:  Soares, M. Bento
;    APPLICANT:  Efstratiadis, Argiris
;    TITLE OF INVENTION:  METHOD FOR CONSTRUCTION OF
;    TITLE OF INVENTION:  NORMALIZED cDNA LIBRARIES
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White, c/o Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,811A
;      FILING DATE:  June 6, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  42840/JPW/AKC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-465-811A-13


  Query Match             0.3%;  Score 19;  DB 1;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgctt 407
```



```
           ||||||||||||||||||
Db      32 AAAAAAAAAAAAAATGCTT 14


RESULT  34
US-08-619-542B-13/c
; Sequence 13, Application US/08619542B
; Patent No. 5830662
;  GENERAL INFORMATION:
;    APPLICANT:  The Trustees of Columbia University in the City
;    APPLICANT:  of New York
;    TITLE OF INVENTION:  METHOD FOR CONSTRUCTION OF NORMALIZED
;    TITLE OF INVENTION:  cDNA LIBRARIES
;    NUMBER OF SEQUENCES:  78
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,542B
;      FILING DATE:  June 21, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  42840-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-619-542B-13


  Query Match             0.3%;  Score 19;  DB 2;  Length 35;
```



```
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgctt 407
           ||||||||||||||||||
Db      32 AAAAAAAAAAAAAATGCTT 14


RESULT  35
US-07-908-245-1
; Sequence 1, Application US/07908245
; Patent No. 5498539
;  GENERAL INFORMATION:
;    APPLICANT:  Harrison, David G.
;    APPLICANT:  Alexander, R. Wayne
;    APPLICANT:  Murphy, T.J.
;    APPLICANT:  Nishida, Ken'ichi
;    TITLE OF INVENTION:  Endothelial Nitric Oxide Synthase
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kilpatrick & Cody
;      STREET:  1100 Peachtree Street, Suite 2800
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  U.S.
;      ZIP:  30309-4530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/908,245
;      FILING DATE:  19920702
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  EMU 111
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404-815-6508
;      TELEFAX:  404-815-6555
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4089 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
```

```
;    ORIGINAL SOURCE:
;      ORGANISM:  Bovine
;      TISSUE TYPE:  Aorta
;      CELL TYPE:  Endothelial
US-07-908-245-1


  Query Match             0.3%;  Score 19;  DB 1;  Length 4089;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     384 taatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||||
Db    4071 TAATGAAAAAAAAAAAAAA 4089


RESULT  36
US-09-331-581-1/c
; Sequence 1, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
;  APPLICANT: TOHDA, Hideki
;  APPLICANT: HAMA, Yuko
;  APPLICANT: KUMAGAI, Hiromicho
;  TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
;  TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
;  TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
;  FILE REFERENCE: 0059-1142-0PCT
;  CURRENT APPLICATION NUMBER: US/09/331,581
;  CURRENT FILING DATE: 1999-06-30
;  EARLIER APPLICATION NUMBER: PCT/JP98/04929
;  EARLIER FILING DATE: 1998-10-30
;  EARLIER APPLICATION NUMBER: JP 9-314608
;  EARLIER FILING DATE: 1997-10-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;  LENGTH: 4748
;  TYPE: DNA
;  ORGANISM: Schizosaccharomyces pombe
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (2810)..(4552)
US-09-331-581-1


  Query Match             0.3%;  Score 19;  DB 3;  Length 4748;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1060 atccgaacggatcaaaagt 1078
```



```
           ||||||||||||||||||||
Db     501 ATCCGAACGGATCAAAAGT 483


RESULT  37
US-09-331-581-3/c
; Sequence 3, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
;  APPLICANT: TOHDA, Hideki
;  APPLICANT: HAMA, Yuko
;  APPLICANT: KUMAGAI, Hiromicho
;  TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
;  TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
;  TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
;  FILE REFERENCE: 0059-1142-0PCT
;  CURRENT APPLICATION NUMBER: US/09/331,581
;  CURRENT FILING DATE: 1999-06-30
;  EARLIER APPLICATION NUMBER: PCT/JP98/04929
;  EARLIER FILING DATE: 1998-10-30
;  EARLIER APPLICATION NUMBER: JP 9-314608
;  EARLIER FILING DATE: 1997-10-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;  LENGTH: 7286
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3


  Query Match             0.3%;  Score 19;  DB 3;  Length 7286;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1060 atccgaacggatcaaaagt 1078
           ||||||||||||||||||||
Db    2016 ATCCGAACGGATCAAAAGT 1998


RESULT  38
US-09-331-581-14/c
; Sequence 14, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
;  APPLICANT: TOHDA, Hideki
;  APPLICANT: HAMA, Yuko
;  APPLICANT: KUMAGAI, Hiromicho
;  TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
```



```
;  TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
;  TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
;  FILE REFERENCE: 0059-1142-0PCT
;  CURRENT APPLICATION NUMBER: US/09/331,581
;  CURRENT FILING DATE: 1999-06-30
;  EARLIER APPLICATION NUMBER: PCT/JP98/04929
;  EARLIER FILING DATE: 1998-10-30
;  EARLIER APPLICATION NUMBER: JP 9-314608
;  EARLIER FILING DATE: 1997-10-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
;  LENGTH: 7938
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-14


  Query Match             0.3%;  Score 19;  DB 3;  Length 7938;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1060 atccgaacggatcaaaagt 1078
           ||||||||||||||||||||
Db    2016 ATCCGAACGGATCAAAAGT 1998


RESULT  39
US-08-870-010A-1/c
; Sequence 1, Application US/08870010A
; Patent No. 5919626
;  GENERAL INFORMATION:
;    APPLICANT:  Shi, Jufang
;    APPLICANT:  Boyce-Jacino, Michael
;    TITLE OF INVENTION:  Covalent Attachment of Unmodified
;    TITLE OF INVENTION:  Nucleic Acids to Silanized Solid Phase Surfaces
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howrey & Simon
;      STREET:  1299 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004-2402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
```



```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/870,010A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sira, Serge
;      REGISTRATION NUMBER:  39,445
;      REFERENCE/DOCKET NUMBER:  04990-0037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202 383-6857
;      TELEFAX:  202 383-6610
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-870-010A-1


  Query Match             0.3%;  Score 18;  DB 2;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaa 399
           ||||||||||||||||||
Db      18 ATTAATGAAAAAAAAAAA 1


RESULT  40
US-09-102-371-1/c
; Sequence 1, Application US/09102371
; Patent No. 6136962
;  GENERAL INFORMATION:
;    APPLICANT:  Shi, Jufang
;    APPLICANT:  Boyce-Jacino, Michael
;    TITLE OF INVENTION:  Covalent Attachment of Unmodified
;    TITLE OF INVENTION:  Nucleic Acids to Silanized Solid Phase Surfaces
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kalow Springut & Bressler LLP
;      STREET:  488 Madison Avenue, 19th Floor
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  PastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/102,371
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Franklin S. Abrams
;       REGISTRATION NUMBER:  43,457
;       REFERENCE/DOCKET NUMBER:  12027-1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212)813-1600
;       TELEFAX:  (212)813-9600
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  37 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-102-371-1


  Query Match           0.3%;  Score 18;  DB 3;  Length 37;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     382 attaatgaaaaaaaaaaa 399
           ||||||||||||||||||
Db      18 ATTAATGAAAAAAAAAAA 1


RESULT  41
US-08-975-902-47
; Sequence 47, Application US/08975902
; Patent No. 5912148
;  GENERAL INFORMATION:
;     APPLICANT:  Perkin-Elmer Corporation, Applied Biosystems Division
;     TITLE OF INVENTION:  Coupled Amplification and Ligation Method
;     NUMBER OF SEQUENCES:  75
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  David J. Weitz, Wilson Sonsini Goodrich & Rosati
;       STREET:  650 Page Mill Road
;       CITY:  Palo Alto
;       STATE:  California
;       COUNTRY:  USA
```



```
;       ZIP:  94304-1050
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch diskette
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  Microsoft Windows 3.1/DOS 5.0
;       SOFTWARE:  Wordperfect for windows 6.0,
;       SOFTWARE:  ASCII (DOS) TEXT format
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/975,902
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/292,686
;       FILING DATE:  19-AUG-94
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  David J. Weitz
;       REGISTRATION NUMBER:  38,362
;       REFERENCE/DOCKET NUMBER:  PELM4215
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 493-9300
;       TELEFAX:  (415) 493-6811
;   INFORMATION FOR SEQ ID NO:  47:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  39 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-975-902-47


  Query Match           0.3%;  Score 18;  DB 2;  Length 39;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgct 406
           ||||||||||||||||||
Db       5 AAAAAAAAAAAAAATGCT 22


RESULT  42
US-09-251-565-47
; Sequence 47, Application US/09251565
; Patent No. 6130073
;  GENERAL INFORMATION:
;     APPLICANT:  Perkin-Elmer Corporation, Applied Biosystems Division
;     TITLE OF INVENTION:  Coupled Amplification and Ligation Method
;     NUMBER OF SEQUENCES:  75
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  David J. Weitz, Wilson Sonsini Goodrich & Rosati
;       STREET:  650 Page Mill Road
;       CITY:  Palo Alto
```



```
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94304-1050
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch diskette
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  Microsoft Windows 98/DOS 5.0
;       SOFTWARE:  Wordperfect for windows 6.0,
;       SOFTWARE:  ASCII (DOS) TEXT format
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/251,565
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/975,902
;       FILING DATE:  19-SEP-96
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/292,686
;       FILING DATE:  19-AUG-94
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  David J. Weitz
;       REGISTRATION NUMBER:  38,362
;       REFERENCE/DOCKET NUMBER:  16842-754
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (650) 493-9300
;       TELEFAX:  (650) 493-6811
;   INFORMATION FOR SEQ ID NO:  47:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  39 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-251-565-47


  Query Match           0.3%;  Score 18;  DB 3;  Length 39;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgct 406
           ||||||||||||||||||
Db       5 AAAAAAAAAAAAAATGCT 22


RESULT  43
US-08-510-032A-3/c
; Sequence 3, Application US/08510032A
; Patent No. 5712126
;  GENERAL INFORMATION:
;     APPLICANT:  Sherman Weissman and Yarindra Prashar
;     TITLE OF INVENTION:  Analysis of Gene Expression By Display of 3'-
```



```
;     TITLE OF INVENTION:  end Restriction Fragments of cDNA
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Yahwak & Associates
;       STREET:  25 Skytop Drive
;       CITY:  Trumbull
;       STATE:  Connecticut
;       COUNTRY:  USA
;       ZIP:  06611
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  floppy disk
;       COMPUTER:  Macintosh
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Microsoft Word 4.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/510,032A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  George M. Yahwak
;       REGISTRATION NUMBER:  26,824
;       REFERENCE/DOCKET NUMBER:  Yale
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (203)268-1951
;       TELEFAX:  (203)268-1951
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  40 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-510-032A-3


  Query Match           0.3%;  Score 18;  DB 1;  Length 40;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgct 406
           ||||||||||||||||||
Db      34 AAAAAAAAAAAAAATGCT 17


RESULT  44
US-08-688-514-3/c
; Sequence 3, Application US/08688514
; Patent No. 6010850
;  GENERAL INFORMATION:
;     APPLICANT:  Sherman Weissman and Yarindra Prashar
;     TITLE OF INVENTION:  Analysis of Gene Expression By Display of 3'-
```

```
;    TITLE OF INVENTION:  end Restriction Fragments of cDNA
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Yahwak & Associates
;      STREET:  25 Skytop Drive
;      CITY:  Trumbull
;      STATE:  Connecticut
;      COUNTRY:  USA
;      ZIP:  06611
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/688,514
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  George M. Yahwak
;      REGISTRATION NUMBER:  26,824
;      REFERENCE/DOCKET NUMBER:  Yale
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (203)268-1951
;      TELEFAX:  (203)268-1951
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-688-514-3


  Query Match             0.3%;  Score 18;  DB 3;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     389 aaaaaaaaaaaaaatgct 406
           ||||||||||||||||||
Db      34 AAAAAAAAAAAAATGCT 17


RESULT  45
US-08-086-410-24
; Sequence 24, Application US/08086410
; Patent No. 5407822
; GENERAL INFORMATION:
;    APPLICANT:  LEPLATOIS, Pascal
;    APPLICANT:  LOISON, Gerard
```



```
;    APPLICANT:  PESSEGUE, Bernard
;    APPLICANT:  SHIRE, David
;    TITLE OF INVENTION:  Artificial promoter for the expression
;    TITLE OF INVENTION:  of proteins in yeast
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  King Street Station,Suite 500,1800 Diagonal
;      STREET:  Road,PO Box 299
;      CITY:  ALEXANDRIA
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/086,410
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/768,083
;      FILING DATE:  02-OCT-1991
;      APPLICATION NUMBER:  FR 89 17467
;      FILING DATE:  29-DEC-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAXE, Bernhard D
;      REGISTRATION NUMBER:  28,665
;      REFERENCE/DOCKET NUMBER:  16781/318
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-9300
;      TELEFAX:  (703) 683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  309 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  ADH2 promoter similar to the natural ADH2
;      CLONE:  promoter
;    FEATURE:
;      NAME/KEY:  misc_binding
;      LOCATION:  150
;      OTHER INFORMATION:  /standard_name= "site SphI"
;    FEATURE:
;      NAME/KEY:  misc_binding
```



```
;      LOCATION:  309
;      OTHER INFORMATION:  /standard_name= "site ClaI"
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  259..270
;    FEATURE:
;      NAME/KEY:  TATA_signal
;      LOCATION:  160..163
US-08-086-410-24


  Query Match             0.3%;  Score 18;  DB 1;  Length 309;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      75 AATGAAAAAAAAAAAAAA 92


RESULT  46
US-07-920-519-28
; Sequence 28, Application US/07920519
; Patent No. 5382518
; GENERAL INFORMATION:
;    APPLICANT:  CAPUT, DANIEL
;    APPLICANT:  FERRARA, PASCUAL
;    APPLICANT:  GUILLEMOT, JEAN-CLAUDE
;    APPLICANT:  KAGHAD, MOURAD
;    APPLICANT:  LEGOUX, RICHARD
;    APPLICANT:  LOISON, GERARD
;    APPLICANT:  LARBRE, ELIZABETH
;    APPLICANT:  LUPKER, JOHANNES
;    APPLICANT:  LEPLATOIS, PASCUAL
;    APPLICANT:  SALOME, MARK
;    TITLE OF INVENTION:  URATE OXIDASE ACTIVITY PROTEIN,
;    TITLE OF INVENTION:  RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,
;    TITLE OF INVENTION:  MICRO-ORGANISMS AND TRANSFORMED CELLS
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```



```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/920,519
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/659,408
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  16781/276 BEDL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  360 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  Plasmid pEMR469 fragment
US-07-920-519-28


  Query Match             0.3%;  Score 18;  DB 1;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      79 AATGAAAAAAAAAAAAAA 96


RESULT  47
US-08-086-410-21
; Sequence 21, Application US/08086410
; Patent No. 5407822
; GENERAL INFORMATION:
;    APPLICANT:  LEPLATOIS, Pascal
;    APPLICANT:  LOISON, Gerard
;    APPLICANT:  PESSEGUE, Bernard
;    APPLICANT:  SHIRE, David
;    TITLE OF INVENTION:  Artificial promoter for the expression
;    TITLE OF INVENTION:  of proteins in yeast
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
```

```
;       STREET:  King Street Station,Suite 500,1800 Diagonal
;       STREET:  Road,PO Box 299
;       CITY:  ALEXANDRIA
;       STATE:  VIRGINIA
;       COUNTRY:  USA
;       ZIP:  22313-0299
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/086,410
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/768,083
;       FILING DATE:  02-OCT-1991
;       APPLICATION NUMBER:  FR 89 17467
;       FILING DATE:  29-DEC-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SAXE, Bernhard D
;       REGISTRATION NUMBER:  28,665
;       REFERENCE/DOCKET NUMBER:  16781/318
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-9300
;       TELEFAX:  (703) 683-4109
;       TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  360 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     IMMEDIATE SOURCE:
;       CLONE:  SmaI-BamHI fragment
;     FEATURE:
;       NAME/KEY:  misc_binding
;       LOCATION:  4
;       OTHER INFORMATION:  /standard_name= "site MluI"
;     FEATURE:
;       NAME/KEY:  misc_binding
;       LOCATION:  154
;       OTHER INFORMATION:  /standard_name= "site SphI"
;     FEATURE:
;       NAME/KEY:  misc_binding
;       LOCATION:  313
;       OTHER INFORMATION:  /standard_name= "site ClaI"
;     FEATURE:
```



```
;       NAME/KEY:  misc_binding
;       LOCATION:  (343^344)
;       OTHER INFORMATION:  /standard_name= "site BamHI"
US-08-086-410-21


  Query Match             0.3%;  Score 18;  DB 1;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      79 AATGAAAAAAAAAAAAAA 96


RESULT  48
US-08-314-586-28
; Sequence 28, Application US/08314586
; Patent No. 5541098
;  GENERAL INFORMATION:
;     APPLICANT:  CAPUT, DANIEL
;     APPLICANT:  FERRARA, PASCUAL
;     APPLICANT:  GUILLEMOT, JEAN-CLAUDE
;     APPLICANT:  KAGHAD, MOURAD
;     APPLICANT:  LEGOUX, RICHARD
;     APPLICANT:  LOISON, GERARD
;     APPLICANT:  LARBRE, ELIZABETH
;     APPLICANT:  LUPKER, JOHANNES
;     APPLICANT:  LEPLATOIS, PASCUAL
;     APPLICANT:  SALOME, MARK
;     APPLICANT:  LAURENT, PATRICK
;     TITLE OF INVENTION:  URATE OXIDASE ACTIVITY PROTEIN,
;     TITLE OF INVENTION:  RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,
;     TITLE OF INVENTION:  MICROORGANISMS AND TRANSFORMED CELLS
;     NUMBER OF SEQUENCES:  40
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, Suite 500
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/314,586
;       FILING DATE:  28-SEP-1994
;       CLASSIFICATION:  435
```



```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/659,408
;       FILING DATE:  25-APR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  16781/509/BEDL
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)836-9300
;       TELEFAX:  (703)683-4109
;       TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  28:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  360 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     IMMEDIATE SOURCE:
;       CLONE:  Plasmid pEMR469 fragment
US-08-314-586-28


  Query Match             0.3%;  Score 18;  DB 1;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      79 AATGAAAAAAAAAAAAAA 96


RESULT  49
US-08-686-878A-22
; Sequence 22, Application US/08686878A
; Patent No. 5708157
;  GENERAL INFORMATION:
;     APPLICANT:  Jacobs, Kenneth
;     APPLICANT:  McCoy, John
;     APPLICANT:  LaVallie, Edward
;     APPLICANT:  Racie, Lisa
;     APPLICANT:  Merberg, David
;     APPLICANT:  Treacy, Maurice
;     APPLICANT:  Evans, Cheryl
;     APPLICANT:  Spaulding, Vikki
;     TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;     TITLE OF INVENTION:  ENCODING THEM
;     NUMBER OF SEQUENCES:  71
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genetics Institute, Inc.
```



```
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/686,878A
;       FILING DATE:
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A.
;       REGISTRATION NUMBER:  32,724
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 498-8224
;       TELEFAX:  (617) 876-5851
;   INFORMATION FOR SEQ ID NO:  22:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  507 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-686-878A-22


  Query Match             0.3%;  Score 18;  DB 1;  Length 507;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     490 AATGAAAAAAAAAAAAAA 507


RESULT  50
US-08-924-759-15
; Sequence 15, Application US/08924759
; Patent No. 5962229
;  GENERAL INFORMATION:
;     APPLICANT:  MCGONIGLE, BRIAN
;     APPLICANT:  O'KEEFE, DANIEL
;     TITLE OF INVENTION:  PLANT GLUTATHIONE-S-TRANSFERASE
;     TITLE OF INVENTION:  ENZYMES
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
```

```
;       STREET:  1007 MARKET STREET
;       CITY:  WILMINGTON
;       STATE:  DELAWARE
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  19898
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  DISKETTE, 3.50 INCH
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;       SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/924,759
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  FLOYD, LINDA AXAMETHY
;       REGISTRATION NUMBER:  33,692
;       REFERENCE/DOCKET NUMBER:  CL-1128
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  302-892-8112
;       TELEFAX:  302-773-0164
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  861 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       TISSUE TYPE:  MAIZE
;     IMMEDIATE SOURCE:
;       CLONE:  CRIN.PK0014.G8
US-08-924-759-15


  Query Match             0.3%;  Score 18;  DB 2;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     829 AATGAAAAAAAAAAAAAA 846


RESULT  51
US-09-248-335-15
; Sequence 15, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
```



```
;  APPLICANT: MCGONIGLE, BRIAN
;  APPLICANT: O'KEEF, DANIEL
;  TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE: CL-1128-A
;  CURRENT APPLICATION NUMBER: US/09/248,335
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: 08/924,759
;  EARLIER FILING DATE: 1997-September-05
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 15
;   LENGTH: 861
;   TYPE: DNA
;   ORGANISM: maize
US-09-248-335-15


  Query Match             0.3%;  Score 18;  DB 3;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     829 aatgaaaaaaaaaaaaaa 846


RESULT  52
US-08-909-965C-9
; Sequence 9, Application US/08909965C
; Patent No. 5936078
;  GENERAL INFORMATION:
;    APPLICANT:  Kuga Tetsro
;    APPLICANT:  Nakagawa Satoshi
;    APPLICANT:  Sakaki yoshiyuki
;    APPLICANT:  Zhao Nanding
;    APPLICANT:  Hashida Hideji
;    TITLE OF INVENTION:  NOVEL DNA, NOVEL POLYPEPTIDE
;    TITLE OF INVENTION:  AND NOVEL ANTIBODY
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FITZPATRICK, CELLA, HARPER AND SCINTO
;      STREET:  277 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10172-0194
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```



```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/909,965C
;      FILING DATE:  August 12, 1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 322745/95
;      APPLICATION NUMBER:  PCT/JP96/03630
;      FILING DATE:  12-No. 5936078-1995
;      FILING DATE:  12-Dec-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lawrence S. Perry
;      REGISTRATION NUMBER:  31865
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-758-2400
;      TELEFAX:  212-758-2982
;      TELEX:  236262
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  882 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  human
;    IMMEDIATE SOURCE:
;      CLONE:  F1180
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  97 to 195
;      IDENTIFICATION METHOD:  by experiment
US-08-909-965C-9


  Query Match             0.3%;  Score 18;  DB 2;  Length 882;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     826 AATGAAAAAAAAAAAAAA 843


RESULT  53
US-08-089-998B-1
; Sequence 1, Application US/08089998B
; Patent No. 5457178
;  GENERAL INFORMATION:
;    APPLICANT:  Jackson et al.
;    TITLE OF INVENTION:  INSECTICIDALLY EFFECTIVE SPIDER
;    TITLE OF INVENTION:  TOXIN
```



```
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz
;      ADDRESSEE:  and No. 5457178ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch disk, 720 Kb
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/089,998B
;      FILING DATE:  07-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  John W. Caldwell
;      REGISTRATION NUMBER:  28,937
;      REFERENCE/DOCKET NUMBER:  FMC-0053
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  940 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
US-08-089-998B-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 940;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     908 AATGAAAAAAAAAAAAAA 925


RESULT  54
US-08-457-272-1
; Sequence 1, Application US/08457272
; Patent No. 5695959
;  GENERAL INFORMATION:
```

```
;     APPLICANT:  Jackson et al.
;     TITLE OF INVENTION:  INSECTICIDALLY EFFECTIVE SPIDER
;     TITLE OF INVENTION:  TOXIN
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz
;       ADDRESSEE:  and No. 5695959ris
;       STREET:  One Liberty Place - 46th Floor
;       CITY:  Philadelphia
;       STATE:  PA
;       COUNTRY:  U.S.A.
;       ZIP:  19103
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch disk, 720 Kb
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/457,272
;       FILING DATE:  Herewith
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/089,998
;       FILING DATE:  07-JUL-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  John W. Caldwell
;       REGISTRATION NUMBER:  28,937
;       REFERENCE/DOCKET NUMBER:  FMC-0073
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  215-568-3100
;       TELEFAX:  215-568-3439
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  940 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
US-08-457-272-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 940;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     908 AATGAAAAAAAAAAAAAA 925


RESULT  55
PCT-US94-07595-1
```



```
; Sequence 1, Application PC/TUS9407595
;  GENERAL INFORMATION:
;    APPLICANT:  Jackson et al.
;     TITLE OF INVENTION:  INSECTICIDALLY EFFECTIVE SPIDER
;     TITLE OF INVENTION:  TOXIN
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FMC Corporation
;       STREET:  1735 Market Street
;       CITY:  Philadelphia
;       STATE:  PA
;       COUNTRY:  U.S.A.
;       ZIP:  19103
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch disk, 720 Kb
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/07595
;       FILING DATE:  Herewith
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/089,998
;       FILING DATE:  07 JUL 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Charles C. Fellows &
;       NAME:  Robert M. Kennedy
;       REGISTRATION NUMBER:  24,261 & 28,026
;       REFERENCE/DOCKET NUMBER:  4797
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  215-299-6970
;       TELEFAX:  215-299-6984
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  940 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
PCT-US94-07595-1


  Query Match             0.3%;  Score 18;  DB 4;  Length 940;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db     908 AATGAAAAAAAAAAAAAA 925
```



```
RESULT  56
US-08-688-988-1
; Sequence 1, Application US/08688988B
; Patent No. 6096545
; GENERAL INFORMATION:
;  APPLICANT: Lefebvre, Daniel D.
;  APPLICANT: Malboobi, Mohammad A.
;  TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
;  FILE REFERENCE: PPL96-03
;  CURRENT APPLICATION NUMBER: US/08/688,988B
;  CURRENT FILING DATE: 1996-07-31
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1250
;   TYPE: DNA
;   ORGANISM: Aribidopsis Thaliana
US-08-688-988-1


  Query Match             0.3%;  Score 18;  DB 3;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1232 aatgaaaaaaaaaaaaaa 1249


RESULT  57
US-09-025-059-2
; Sequence 2, Application US/09025059
; Patent No. 6075136
;  GENERAL INFORMATION:
;     APPLICANT:  Tang, Y. Tom
;     APPLICANT:  Corley, Neil C.
;     APPLICANT:  Guegler, Karl J.
;     TITLE OF INVENTION:  PROSTATE-ASSOCIATED SERINE PROTEASE
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;       STREET:  3174 Porter Dr.
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
```



```
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/025,059
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0481 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-855-0555
;       TELEFAX:  650-845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1314 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  LUNGTUT10
;       CLONE:  2723646
US-09-025-059-2


  Query Match             0.3%;  Score 18;  DB 3;  Length 1314;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1281 AATGAAAAAAAAAAAAAA 1298


RESULT  58
US-08-270-076A-1/c
; Sequence 1, Application US/08270076A
; Patent No. 5667986
;  GENERAL INFORMATION:
;     APPLICANT:  Sleep, Darrell
;     APPLICANT:  Goodey, Andrew R
;     APPLICANT:  Vakeria, Diana
;     TITLE OF INVENTION:  Yeast Promoter
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  The BOC Group, Inc.
;       STREET:  100 Mountain Avenue, Murray Hill
;       CITY:  New Providence
;       STATE:  New Jersey
```

```
;       COUNTRY:  USA
;       ZIP:  07974
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/270,076A
;       FILING DATE:  01-JUL-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 8923521.2
;       FILING DATE:  18-OCT-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/597,687
;       FILING DATE:  16-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/925,286
;       FILING DATE:  04-AUG-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Swope, R. Hain
;       REGISTRATION NUMBER:  24864
;       REFERENCE/DOCKET NUMBER:  92H834-3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  908/771-6292
;       TELEFAX:  908/771-6159
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1357 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Saccharomyces cerevisiae
;     FEATURE:
;       NAME/KEY:  promoter
;       LOCATION:  42..52
;       OTHER INFORMATION:  /function= "RNA POLIII
;       OTHER INFORMATION:  promoter box A"
;     FEATURE:
;       NAME/KEY:  promoter
;       LOCATION:  86..96
;       OTHER INFORMATION:  /function= "RNA POLIII
;       OTHER INFORMATION:  Promoter box B"
;     FEATURE:
;       NAME/KEY:  promoter
;       LOCATION:  113..118
```



```
;       OTHER INFORMATION:  /function= "RNA POLIII
;       OTHER INFORMATION:  Terminator"
;     FEATURE:
;       NAME/KEY:  protein_bind
;       LOCATION:  1091..1103
;       OTHER INFORMATION:  /bound_moiety=
;       OTHER INFORMATION:  "RAPI/GRFI/TUFI"
;     FEATURE:
;       NAME/KEY:  protein_bind
;       LOCATION:  1106..1118
;       OTHER INFORMATION:  /bound_moiety=
;       OTHER INFORMATION:  "RAPI/GRFI/TUFI"
;     FEATURE:
;       NAME/KEY:  misc_signal
;       LOCATION:  1176..1241
;       OTHER INFORMATION:  /function= "Pyrimidine (CT)
;       OTHER INFORMATION:  block"
;     FEATURE:
;       NAME/KEY:  TATA_signal
;       LOCATION:  1326..1335
;     FEATURE:
;       NAME/KEY:  misc_signal
;       LOCATION:  1246..1283
;       OTHER INFORMATION:  /function= "Pyrimidine (CT)
;       OTHER INFORMATION:  block"
;     FEATURE:
;       NAME/KEY:  misc_signal
;       LOCATION:  1295..1298
;       OTHER INFORMATION:  /function= "CAAG box"
;     FEATURE:
;       NAME/KEY:  misc_signal
;       LOCATION:  1302..1306
;       OTHER INFORMATION:  /function= "CAATT box"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1031..1036
;       OTHER INFORMATION:  /function= "PstI restriction site"
US-08-270-076A-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 1357;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    387 tgaaaaaaaaaaaaaatg 404
          ||||||||||||||||||
Db    436 TGAAAAAAAAAAAAAATG 419


RESULT  59
US-08-394-189B-3
```



```
; Sequence 3, Application US/08394189B
; Patent No. 5952301
;  GENERAL INFORMATION:
;    APPLICANT:  Horvitz, Robert
;    APPLICANT:  Yuan, Junying
;    APPLICANT:  Shaham, Shai
;    TITLE OF INVENTION:  RELATEDNESS OF HUMAN INTERLEUKIN-1
;    TITLE OF INVENTION:  BETA CONVERTASE GENE TO A C. ELEGANS CELL DEATH
;    TITLE OF INVENTION:  GENE, INHIBITORY PORTIONS OF THESE GENES AND...
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/394,189B
;      FILING DATE:  24-FEB-2005
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/282,211
;      FILING DATE:  12-JUL-1994
;      APPLICATION NUMBER:  07/984,182
;      FILING DATE:  20-NOV-1992
;      APPLICATION NUMBER:  07/897,788
;      FILING DATE:  12-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:  39,109
;      REFERENCE/DOCKET NUMBER:  01997/211001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-428-0200
;      TELEFAX:  617-428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1373 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  18...1229
```



```
;       OTHER INFORMATION:  /product= "human interleukin-1 beta
;       OTHER INFORMATION:  convertase"
US-08-394-189B-3


  Query Match             0.3%;  Score 18;  DB 2;  Length 1373;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1352 AATGAAAAAAAAAAAAAA 1369


RESULT  60
PCT-US93-05705-3
; Sequence 3, Application PC/TUS9305705
;  GENERAL INFORMATION:
;    APPLICANT:  Massachusetts Institute of Technology
;    TITLE OF INVENTION:  Inhibitors of Ced-3 and Related Proteins
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Massachusetts Institute of Technology
;      STREET:  77 Massachusetts Avenue
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02139
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  diskette
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05705
;      FILING DATE:  19930714
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1373 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  18..1232
;      OTHER INFORMATION:  /product= "human interleukin-1
;      OTHER INFORMATION:  convertase"
PCT-US93-05705-3


  Query Match             0.3%;  Score 18;  DB 4;  Length 1373;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1352 AATGAAAAAAAAAAAAAA 1369


RESULT  61
US-08-270-076A-2/c
; Sequence 2, Application US/08270076A
; Patent No. 5667986
;  GENERAL INFORMATION:
;    APPLICANT:  Sleep, Darrell
;    APPLICANT:  Goodey, Andrew R
;    APPLICANT:  Vakeria, Diana
;    TITLE OF INVENTION:  Yeast Promoter
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The BOC Group, Inc.
;      STREET:  100 Mountain Avenue, Murray Hill
;      CITY:  New Providence
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07974
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/270,076A
;      FILING DATE:  01-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8923521.2
;      FILING DATE:  18-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/597,687
;      FILING DATE:  16-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/925,286
;      FILING DATE:  04-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Swope, R. Hain
;      REGISTRATION NUMBER:  24864
;      REFERENCE/DOCKET NUMBER:  92H834-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908/771-6292
;      TELEFAX:  908/771-6159
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1483 base pairs
```



```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  123..124
;      OTHER INFORMATION:  /product= "SatII restriction site"
US-08-270-076A-2


  Query Match             0.3%;  Score 18;  DB 1;  Length 1483;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    387 tgaaaaaaaaaaaaaatg 404
          ||||||||||||||||||
Db    559 TGAAAAAAAAAAAAAATG 542


RESULT  62
US-08-582-740-69
; Sequence 69, Application US/08582740
; Patent No. 6037324
;  GENERAL INFORMATION:
;    APPLICANT:  Schwender, Charles F.
;    APPLICANT:  Shroff, Hitesh N.
;    TITLE OF INVENTION:  Inhibitors of MADCAM-1-Mediated
;    TITLE OF INVENTION:  Interactions and Methods of Use Therefor
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/582,740
;      FILING DATE:  04-JAN-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  LKS95-12
;    TELECOMMUNICATION INFORMATION:
```



```
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  69:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1539 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1146
US-08-582-740-69


  Query Match             0.3%;  Score 18;  DB 3;  Length 1539;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1521 AATGAAAAAAAAAAAAAA 1538


RESULT  63
US-08-597-545-4
; Sequence 4, Application US/08597545
; Patent No. 5580738
;  GENERAL INFORMATION:
;    APPLICANT:  LABORDA, Jorge
;    TITLE OF INVENTION:  Delta-Like Gene Expressed In
;    TITLE OF INVENTION:  Neuroendocrine Tumors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/597,545
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/989,537
;      FILING DATE:  11-DEC-1992
```



```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/166 NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1573 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  Mouse Dlk
US-08-597-545-4


  Query Match             0.3%;  Score 18;  DB 1;  Length 1573;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1551 AATGAAAAAAAAAAAAAA 1568


RESULT  64
US-08-457-135-4
; Sequence 4, Application US/08457135
; Patent No. 5644031
;  GENERAL INFORMATION:
;    APPLICANT:  LABORDA, Jorge
;    TITLE OF INVENTION:  Delta-Like Gene Expressed In
;    TITLE OF INVENTION:  Neuroendocrine Tumors
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,135
;      FILING DATE:  01-JUN-1995
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/989,537
;      FILING DATE:  11-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/304/NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;      LENGTH:  1573 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      STRAIN:  Mouse Dlk
US-08-457-135-4


  Query Match             0.3%;  Score 18;  DB 1;  Length 1573;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1551 AATGAAAAAAAAAAAAAA 1568


RESULT  65
US-08-203-716-1
; Sequence 1, Application US/08203716
; Patent No. 5416013
;  GENERAL INFORMATION:
;    APPLICANT:  BLACK, ROY A
;    APPLICANT:  SLEATH, PAUL R
;    APPLICANT:  KRONHEIM, SHIRLEY R
;    TITLE OF INVENTION:  INTERLEUKIN 1B PROTEASE AND
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DRESSLER,GOLDSMITH, SHORE, SUTKER &
;      STREET:  180 N. STETSON
;      CITY:  CHICAGO
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```



```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/203,716
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/750,644
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KATZ, MARTIN L.
;      REGISTRATION NUMBER:  25011
;      REFERENCE/DOCKET NUMBER:  IMMUNEX2108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  3126165400
;      TELEFAX:  3126165460
;      TELEX:  9102211206
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1659 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-203-716-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 1659;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1638 AATGAAAAAAAAAAAAAA 1655


RESULT  66
US-08-440-179-1
; Sequence 1, Application US/08440179
; Patent No. 5756465
;  GENERAL INFORMATION:
;    APPLICANT:  BLACK, ROY A
;    APPLICANT:  SLEATH, PAUL R
;    APPLICANT:  KRONHEIM, SHIRLEY R
;    TITLE OF INVENTION:  INTERLEUKIN 1B PROTEASE AND
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DRESSLER,GOLDSMITH, SHORE, SUTKER &
;      ADDRESSEE:  MILNAMOW
;      STREET:  180 N. STETSON
;      CITY:  CHICAGO
;      STATE:  IL
```



```
;      COUNTRY:  USA
;      ZIP:  60601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/440,179
;      FILING DATE:  12-MAY-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/750,644
;      FILING DATE:  30-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KATZ, MARTIN L.
;      REGISTRATION NUMBER:  25011
;      REFERENCE/DOCKET NUMBER:  IMMUNEX2108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  3126165400
;      TELEFAX:  3126165460
;      TELEX:  9102211206
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1659 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-440-179-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 1659;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1638 AATGAAAAAAAAAAAAAA 1655


RESULT  67
US-09-039-657-1
; Sequence 1, Application US/09039657
; Patent No. 6136787
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Roy A.
;    APPLICANT:  Sleath, Paul R.
;    APPLICANT:  Kronheim, Shirley R.
;    TITLE OF INVENTION:  INTERLEUKIN 1B PROTEASE AND INTERLEUKIN
;    TITLE OF INVENTION:  1B PROTEASE INHIBITORS
;    NUMBER OF SEQUENCES:  24
```



```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/039,657
;      FILING DATE:  16-MAR-1998
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley Jr., James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  VPI/SW002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1659 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-09-039-657-1


  Query Match             0.3%;  Score 18;  DB 3;  Length 1659;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1638 AATGAAAAAAAAAAAAAA 1655


RESULT  68
PCT-US91-02339-2
; Sequence 2, Application PC/TUS9102339
;  GENERAL INFORMATION:
;    APPLICANT:  BLACK, ROY
;    APPLICANT:  SLEATH, PAUL
;    APPLICANT:  KRONHEIM, SHIRLEY
;    TITLE OF INVENTION:  INTERLEUKIN 1 B PROTEASE
;    NUMBER OF SEQUENCES:  4
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  IMMUNEX CORPORATION
;       STREET:  51 UNIVERSITY STREET
;       CITY:  SEATTLE
;       STATE:  WASHINGTON
;       COUNTRY:  USA
;       ZIP:  98101
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.24
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US91/02339
;       FILING DATE:  19910404
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  OSTER, JEFFREY B.
;       REGISTRATION NUMBER:  32,585
;       REFERENCE/DOCKET NUMBER:  0401WO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  2065870430
;       TELEFAX:  2065870606
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1659 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  N
;     ANTI-SENSE:  N
;     FRAGMENT TYPE:  N-terminal
;     ORIGINAL SOURCE:
;       ORGANISM:  INTERLEUKIN 1B PROTEASE
;     IMMEDIATE SOURCE:
;       CLONE:  INTERLEUKIN 1B PROTEASE
;     PUBLICATION INFORMATION:
;       AUTHORS:  BLACK , ROY
;       AUTHORS:  SLEATH, PAUL
;       AUTHORS:  KRONHEIM, SHIRLEY
PCT-US91-02339-2


  Query Match             0.3%;  Score 18;  DB 4;  Length 1659;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1638 AATGAAAAAAAAAAAAAA 1655
```



```
RESULT  69
US-08-698-551-9
; Sequence 9, Application US/08698551
; Patent No. 5712381
;  GENERAL INFORMATION:
;     APPLICANT:  Lin, Lih-Ling
;     APPLICANT:  Chen, Jennifer H.
;     APPLICANT:  Schievella, Andrea
;     APPLICANT:  Graham, James
;     TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;     TITLE OF INVENTION:  PROTEINS AND INHIBITORS OF LIGAND BINDING
;     NUMBER OF SEQUENCES:  18
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genetics Institute, Inc.
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/698,551
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A.
;       REGISTRATION NUMBER:  32,724
;       REFERENCE/DOCKET NUMBER:  GI5232D
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 498-8224
;       TELEFAX:  (617) 876-5851
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1694 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  2..931
US-08-698-551-9
```



```
  Query Match             0.3%;  Score 18;  DB 1;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1672 AATGAAAAAAAAAAAAAA 1689


RESULT  70
US-08-602-228-9
; Sequence 9, Application US/08602228
; Patent No. 5843675
;  GENERAL INFORMATION:
;     APPLICANT:  Lin, Lih-Ling
;     APPLICANT:  Chen, Jennifer H.
;     APPLICANT:  Schievella, Andrea
;     APPLICANT:  Graham, James
;     TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;     TITLE OF INVENTION:  PROTEINS AND INHIBITORS OF LIGAND BINDING
;     NUMBER OF SEQUENCES:  16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genetics Institute, Inc.
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/602,228
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A.
;       REGISTRATION NUMBER:  32,724
;       REFERENCE/DOCKET NUMBER:  GI5232C
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 498-8224
;       TELEFAX:  (617) 876-5851
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1694 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
```



```
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  2..931
US-08-602-228-9


  Query Match             0.3%;  Score 18;  DB 2;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1672 AATGAAAAAAAAAAAAAA 1689


RESULT  71
US-08-494-440B-9
; Sequence 9, Application US/08494440B
; Patent No. 5849501
;  GENERAL INFORMATION:
;     APPLICANT:  Lin, Lih-Ling
;     APPLICANT:  Chen, Jennifer H.
;     APPLICANT:  Schievella, Andrea
;     TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;     TITLE OF INVENTION:  PROTEINS AND INHIBITORS OF LIGAND BINDING
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genetics Institute, Inc.
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/494,440B
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A.
;       REGISTRATION NUMBER:  32,724
;       REFERENCE/DOCKET NUMBER:  GI5232A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 498-8224
;       TELEFAX:  (617) 876-5851
;   INFORMATION FOR SEQ ID NO:  9:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1694 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..931
US-08-494-440B-9


  Query Match              0.3%;  Score 18;  DB 2;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1672 AATGAAAAAAAAAAAAAA 1689


RESULT  72
US-08-533-901B-9
; Sequence 9, Application US/08533901B
; Patent No. 5852173
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Chen, Jennifer H.
;    APPLICANT:  Schievella, Andrea
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;    TITLE OF INVENTION:  PROTEINS AND INHIBITORS OF LIGAND BINDING
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,901B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```



```
;      NAME:  Brown, Scott A,
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5232
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1694 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..931
US-08-533-901B-9


  Query Match              0.3%;  Score 18;  DB 2;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1672 AATGAAAAAAAAAAAAAA 1689


RESULT  73
US-08-839-032A-9
; Sequence 9, Application US/08839032A
; Patent No. 5891675
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Chen, Jennifer H.
;    APPLICANT:  Schievella, Andrea
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```



```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/839,032A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A,
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5232DDIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1694 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..931
US-08-839-032A-9


  Query Match              0.3%;  Score 18;  DB 2;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1672 AATGAAAAAAAAAAAAAA 1689


RESULT  74
US-08-839-031A-9
; Sequence 9, Application US/08839031A
; Patent No. 5948638
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Chen, Jennifer H.
;    APPLICANT:  Schievella, Andrea
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
```



```
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/839,031A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sprunger, Suzanne A.,
;      REGISTRATION NUMBER:  41,323
;      REFERENCE/DOCKET NUMBER:  GI5232BDIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8284
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1694 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..931
US-08-839-031A-9


  Query Match              0.3%;  Score 18;  DB 2;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1672 AATGAAAAAAAAAAAAAA 1689


RESULT  75
PCT-US95-12724-9
; Sequence 9, Application PC/TUS9512724
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Chen, Jennifer H.
;    APPLICANT:  Schievella, Andrea
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
```

```
;     TITLE OF INVENTION:  PROTEINS AND INHIBITORS OF LIGAND-BINDING
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Genetics Institute, Inc.
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/12724
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A,
;       REGISTRATION NUMBER:  32,724
;       REFERENCE/DOCKET NUMBER:  GI5232B
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 498-8224
;       TELEFAX:  (617) 876-5851
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1694 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  2..931
PCT-US95-12724-9


  Query Match             0.3%;  Score 18;  DB 4;  Length 1694;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1672 AATGAAAAAAAAAAAAAA 1689


RESULT  76
US-07-841-646-1
; Sequence 1, Application US/07841646
```



```
; Patent No. 5266683
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/841,646
;      FILING DATE:  19920221
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
```



```
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
US-07-841-646-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
```



```
    Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  77
US-07-901-703-1
; Sequence 1, Application US/07901703
; Patent No. 5344654
;  GENERAL INFORMATION:
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZAKAYNAK, ENGIN
;    TITLE OF INVENTION:  PROSTHETIC DEVICES HAVING ENHANCED
;    TITLE OF INVENTION:  OSTEOGENIC PROPERTIES
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  EXCHANGE PLACE, 53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/901,703
;      FILING DATE:  19920616
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER  ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  STK-057
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
```

```
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
US-07-901-703-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  78
US-08-147-023-1
; Sequence 1, Application US/08147023
; Patent No. 5488045
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,023
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
```



```
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
```



```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
US-08-147-023-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  79
US-08-206-864-1
; Sequence 1, Application US/08206864
; Patent No. 5610021
;  GENERAL INFORMATION:
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  JONES, WILLIAM K
;    APPLICANT:  TUCKER, RONALD F
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR
;    TITLE OF INVENTION:  RECOMBINANT OSTEOGENIC PROTEIN PRODUCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR/CREATIVE BIOMOLECULES,
;      ADDRESSEE:  INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
```



```
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/206,864
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/027,070
;      FILING DATE:  04-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/841,646
;      FILING DATE:  21-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-096
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
US-08-206-864-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  80
US-08-278-729A-16
; Sequence 16, Application US/08278729A
; Patent No. 5650276
;  GENERAL INFORMATION:
;    APPLICANT:  SMART, JOHN
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN SCREENING METHOD
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/278,729A
;      FILING DATE:  20-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER Esq., EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-058CPFW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
```



```
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-278-729A-16


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  81
US-08-480-528A-3
; Sequence 3, Application US/08480528A
; Patent No. 5652118
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,528A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-076FW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
```



```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP1-PP"
;      OTHER INFORMATION:  /note= "hOP1 cDNA"
US-08-480-528A-3


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  82
US-08-479-666-3
; Sequence 3, Application US/08479666
; Patent No. 5652337
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```



```
;      APPLICATION NUMBER:  US/08/479,666
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-076DV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP1-PP"
;      OTHER INFORMATION:  /note= "hOP1 cDNA"
US-08-479-666-3


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  83
US-08-155-343A-16
; Sequence 16, Application US/08155343A
; Patent No. 5656593
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  MORPHOGENIC-INDUCED PERIODONTAL TISSUE
;    TITLE OF INVENTION:  REGENERATION.
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
```

```
;       CITY:  HOPKINTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  01748
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/155,343A
;       FILING DATE:  15-NOV-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  FENTON Esq., GILLIAN M.
;       REGISTRATION NUMBER:  36,508
;       REFERENCE/DOCKET NUMBER:  CRP-067FW
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248-7560
;       TELEFAX:  (617) 248-7100
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..1341
;       OTHER INFORMATION:  /product= "HOP-1"
US-08-155-343A-16


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   385 aatgaaaaaaaaaaaaaa 402
         ||||||||||||||||||
Db  1801 AATGAAAAAAAAAAAAAA 1818


RESULT  84
US-08-406-672-16
; Sequence 16, Application US/08406672
; Patent No. 5674844
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  OPPERMANN, HERMANN
```



```
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  TREATMENT TO PREVENT LOSS OF AND/OR
;    TITLE OF INVENTION:  INCREASE BONE MASS IN METABOLIC BONE DISEASES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/406,672
;      FILING DATE:  20-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 752,857
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 667,274
;      FILING DATE:  11-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-060CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-406-672-16


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
```



```
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   385 aatgaaaaaaaaaaaaaa 402
         ||||||||||||||||||
Db  1801 AATGAAAAAAAAAAAAAA 1818


RESULT  85
US-08-643-563A-16
; Sequence 16, Application US/08643563A
; Patent No. 5707810
;  GENERAL INFORMATION:
;    APPLICANT:  SMART, JOHN
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN SCREENING METHOD
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/643,563A
;      FILING DATE:  06-MAY-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  THOMEY Esq., MICHAEL J.
;      REGISTRATION NUMBER:  38,349
;      REFERENCE/DOCKET NUMBER:  CRP-058CN2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
```



```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-643-563A-16


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   385 aatgaaaaaaaaaaaaaa 402
         ||||||||||||||||||
Db  1801 AATGAAAAAAAAAAAAAA 1818


RESULT  86
US-08-447-570-1
; Sequence 1, Application US/08447570
; Patent No. 5714589
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,570
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US 660,162
;       FILING DATE:  22-FEB-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,988
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,849
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 616,374
;       FILING DATE:  21-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 600,024
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 599,543
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 579,865
;       FILING DATE:  07-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 569,920
;       FILING DATE:  20-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 483,913
;       FILING DATE:  22-FEB-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 422,613
;       FILING DATE:  17-OCT-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 315,342
;       FILING DATE:  23-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 232,630
;       FILING DATE:  15-AUG-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 179,460
;       FILING DATE:  08-APR-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP-001CP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
```



```
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  HOMO SAPIENS
;       TISSUE TYPE:  HIPPOCAMPUS
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..1341
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "BP1"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "OP1"
US-08-447-570-1


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  87
US-08-643-763A-16
; Sequence 16, Application US/08643763A
; Patent No. 5733878
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  MORPHOGENIC-INDUCED PERIODONTAL TISSUE
;    TITLE OF INVENTION:  REGENERATION.
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```



```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/643,763A
;      FILING DATE:  06-MAY-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-067CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-643-763A-16


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  88
US-08-462-623-16
; Sequence 16, Application US/08462623
; Patent No. 5739107
;  GENERAL INFORMATION:
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  CHARETTE, MARC F.
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  SMART, JOHN E.
;    TITLE OF INVENTION:  MORPHOGEN TREATMENT OF GASTROINTESTINAL
;    TITLE OF INVENTION:  ULCERS.
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
```



```
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,623
;      FILING DATE:
;      CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/445,882
;      FILING DATE:  22-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-074CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-462-623-16


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  89
US-08-451-953A-16
; Sequence 16, Application US/08451953A
```

```
; Patent No. 5741641
;  GENERAL INFORMATION:
;    APPLICANT:  SMART, JOHN
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN SCREENING METHOD
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/451,953A
;      FILING DATE:  26-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER Esq., EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-058CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-451-953A-16


  Query Match             0.3%;  Score 18;  DB 1;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT  90
US-08-459-346-1
; Sequence 1, Application US/08459346
; Patent No. 5834179
;  GENERAL INFORMATION:
;    APPLICANT:  JONES, WILLIAM K
;    APPLICANT:  TUCKER, RONALD F
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    TITLE OF INVENTION:  NOVEL MORPHOGENIC PROTEIN COMPOSITIONS
;    TITLE OF INVENTION:  OF MATTER
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR/CREATIVE BIOMOLECULES,
;      ADDRESSEE:  INC.
;      STREET:  35 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,346
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/029,335
;      FILING DATE:  04-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,091
;      FILING DATE:  03-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/946,235
;      FILING DATE:  16-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,336
;      FILING DATE:  08-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/923,780
```



```
;      FILING DATE:  31-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-081CP
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
US-08-459-346-1


  Query Match             0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT  91
US-08-445-468A-16
; Sequence 16, Application US/08445468A
; Patent No. 5849686
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED LIVER REGENERATION
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
```



```
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/445,468A
;      FILING DATE:  22-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-072FW2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-445-468A-16


  Query Match             0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT  92
US-08-901-200A-3
; Sequence 3, Application US/08901200A
; Patent No. 5854071
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
```

```
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  RUEGER, DAVID C.
;     APPLICANT:  PANG, ROY H.L.
;     APPLICANT:  COHEN, CHARLES M.
;     TITLE OF INVENTION:  OPJ-INDUCED MORPHOGENESIS
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;       STREET:  45 SOUTH STREET
;       CITY:  HOPKINTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  01748
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/901,200A
;       FILING DATE:  28-JUL-1997
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MEYERS, THOMAS C.
;       REGISTRATION NUMBER:  36,989
;       REFERENCE/DOCKET NUMBER:  CRP-076DV2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (508) 435-9001
;       TELEFAX:  (508) 435-6951
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..1341
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "hOP1-PP"
;       OTHER INFORMATION:  /note= "hOP1 cDNA"
US-08-901-200A-3


  Query Match             0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
```



```
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  93
US-08-481-337A-9
; Sequence 9, Application US/08481337A
; Patent No. 5863738
;  GENERAL INFORMATION:
;     APPLICANT:  TEN DIJKE, Peter
;     APPLICANT:  HELDIN, Carl-Henrik
;     APPLICANT:  MIYAZONO, Kohei
;     APPLICANT:  SAMPATH, Kuber T.
;     TITLE OF INVENTION:  Morphogenic Protein-Specific Cell
;     TITLE OF INVENTION:  Surface Receptors and Uses Therefor
;     NUMBER OF SEQUENCES:  18
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Testa, Hurwitz & Thibeault
;       STREET:  125 High St.
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/481,337A
;       FILING DATE:  02-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MEYERS, Thomas C.
;       REGISTRATION NUMBER:  36,989
;       REFERENCE/DOCKET NUMBER:  CRP-097CP2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 248-7000
;       TELEFAX:  (617) 248-7100
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..1341
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "OP1"
```



```
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "OP1"
US-08-481-337A-9


  Query Match             0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  94
US-08-449-700-1
; Sequence 1, Application US/08449700
; Patent No. 5863758
;  GENERAL INFORMATION:
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  OZKAYNAK, ENGIN
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  RUEGER, DAVID C.
;     APPLICANT:  PANG, ROY H.L.
;     TITLE OF INVENTION:  OSTEOGENIC DEVICES
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  53 STATE STREET
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  U.S.A.
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/449,700
;       FILING DATE:  21-FEB-1992
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 810,560
;       FILING DATE:  20-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 827,052
;       FILING DATE:  28-JAN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 660,162
;       FILING DATE:  22-FEB-1991
```



```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,988
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,849
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 616,374
;       FILING DATE:  21-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 600,024
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 599,543
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 579,865
;       FILING DATE:  07-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 569,920
;       FILING DATE:  20-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 483,913
;       FILING DATE:  22-FEB-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 422,613
;       FILING DATE:  17-OCT-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 315,342
;       FILING DATE:  23-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 232,630
;       FILING DATE:  15-AUG-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 179,460
;       FILING DATE:  08-APR-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP-001CP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
```

```
;  ANTI-SENSE:  NO
;  ORIGINAL SOURCE:
;    ORGANISM:  HOMO SAPIENS
;    TISSUE TYPE:  HIPPOCAMPUS
;  FEATURE:
;    NAME/KEY:  CDS
;    LOCATION:  49..1341
;    IDENTIFICATION METHOD:  experimental
;    OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;    OTHER INFORMATION:  /product= "OP1"
;    OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    OTHER INFORMATION:  /standard_name= "OP1"
US-08-449-700-1


  Query Match           0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  95
US-08-449-699A-1
; Sequence 1, Application US/08449699A
; Patent No. 5958441
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  ANTIBODIES TO OSTEOGENIC PROTEINS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/449,699A
;      FILING DATE:  24-MAY-1995
```



```
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/147,023
;      FILING DATE:  01-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  STK-001CP6CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
US-08-449-699A-1


  Query Match           0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT  96
US-08-696-268B-3
; Sequence 3, Application US/08696268B
; Patent No. 5968752
;  GENERAL INFORMATION:
;    APPLICANT:  ICHIJO, HIDENORI
;    APPLICANT:  NISHITOH, HIDEKI
;    APPLICANT:  SAMPATH, KUBER T.
```



```
;    TITLE OF INVENTION:  NOVEL SIGNALLING RECEPTOR FOR
;    TITLE OF INVENTION:  MORPHOGENIC PROTEINS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,268B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, Thomas C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  CRP-117
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /function= "Osteogenic Protein"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /standard_name= "OP1"
US-08-696-268B-3


  Query Match           0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818
```



```
RESULT  97
US-08-461-397A-16
; Sequence 16, Application US/08461397A
; Patent No. 5972884
;  GENERAL INFORMATION:
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  CHARETTE, MARC F.
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  SMART, JOHN E.
;    TITLE OF INVENTION:  MORPHOGEN TREATMENT FOR LIMITING
;    TITLE OF INVENTION:  PROLIFERATION OF EPITHELIAL CELLS.
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,397A
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-074PW2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-461-397A-16
```

```
  Query Match             0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      385 aatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||
Db     1801 AATGAAAAAAAAAAAAAA 1818


RESULT 98
US-08-912-088-16
; Sequence 16, Application US/08912088
; Patent No. 5994131
; GENERAL INFORMATION:
;   APPLICANT:  SMART, JOHN
;   APPLICANT:  OPPERMANN, HERMAN
;   APPLICANT:  OZKAYNAK, ENGIN
;   APPLICANT:  KUBERASAMPATH, THANGAVEL
;   APPLICANT:  RUEGER, DAVID C.
;   APPLICANT:  PANG, ROY H.L.
;   APPLICANT:  COHEN, CHARLES M.
;   TITLE OF INVENTION:  MORPHOGENIC PROTEIN SCREENING METHOD
;   NUMBER OF SEQUENCES:  33
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;     STREET:  45 SOUTH STREET
;     CITY:  HOPKINTON
;     STATE:  MA
;     COUNTRY:  USA
;     ZIP:  01748
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/912,088
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/278,729
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  PITCHER Esq., EDMUND R.
;     REGISTRATION NUMBER:  27,829
;     REFERENCE/DOCKET NUMBER:  CRP-058CPFW
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (508) 435-9001
;     TELEFAX:  (508) 435-6951
```



```
; INFORMATION FOR SEQ ID NO:  16:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1822 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  49..1341
;     OTHER INFORMATION:  /product= "HOP-1"
US-08-912-088-16


  Query Match             0.3%;  Score 18;  DB 2;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      385 aatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||
Db     1801 AATGAAAAAAAAAAAAAA 1818


RESULT 99
US-08-278-730A-16
; Sequence 16, Application US/08278730A
; Patent No. 6022853
; GENERAL INFORMATION:
;   APPLICANT:  KUBERASAMPATH, THANGAVEL
;   APPLICANT:  COHEN, CHARLES M.
;   APPLICANT:  RUEGER, DAVID C.
;   APPLICANT:  OPPERMANN, HERMAN
;   APPLICANT:  PANG, ROY H.L.
;   TITLE OF INVENTION:  MORPHOGENIC-ENRICHED DIETARY COMPOSITION
;   NUMBER OF SEQUENCES:  33
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;     STREET:  45 SOUTH STREET
;     CITY:  HOPKINTON
;     STATE:  MA
;     COUNTRY:  USA
;     ZIP:  01748
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/278,730A
;     FILING DATE:  20-JULY-1994
;     CLASSIFICATION:  435
```



```
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  FENTON Esq., GILLIAN M.
;     REGISTRATION NUMBER:  36,508
;     REFERENCE/DOCKET NUMBER:  CRP-071FW
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (617) 248-7560
;     TELEFAX:  (617) 248-7100
; INFORMATION FOR SEQ ID NO:  16:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1822 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  49..1341
;     OTHER INFORMATION:  /product= "HOP-1"
US-08-278-730A-16


  Query Match             0.3%;  Score 18;  DB 3;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      385 aatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||
Db     1801 AATGAAAAAAAAAAAAAA 1818


RESULT 100
US-08-458-811-1
; Sequence 1, Application US/08458811
; Patent No. 6027743
; GENERAL INFORMATION:
;   APPLICANT:  KHOURI, ROGER K.
;   APPLICANT:  SAMPATH, KUBER T.
;   APPLICANT:  RUEGER, DAVID C.
;   TITLE OF INVENTION:  MANUFACTURE OF AUTOGENOUS REPLACEMENT
;   TITLE OF INVENTION:  BODY PARTS
;   NUMBER OF SEQUENCES:  3
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;     STREET:  53 STATE STREET
;     CITY:  BOSTON
;     STATE:  MASSACHUSETTS
;     COUNTRY:  U.S.A.
;     ZIP:  02109
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
```



```
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/458,811
;     FILING DATE:
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  KELLEY, ROBIN D.
;     REGISTRATION NUMBER:  34,637
;     REFERENCE/DOCKET NUMBER:  CRP-108
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  617/248-7000
;     TELEFAX:  617/248-7100
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1822 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   ORIGINAL SOURCE:
;     ORGANISM:  HOMO SAPIENS
;     TISSUE TYPE:  HIPPOCAMPUS
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  49..1341
;     IDENTIFICATION METHOD:  experimental
;     OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;     OTHER INFORMATION:  /product= "OP1"
;     OTHER INFORMATION:  /evidence= EXPERIMENTAL
;     OTHER INFORMATION:  /standard_name= "OP1"
US-08-458-811-1


  Query Match             0.3%;  Score 18;  DB 3;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      385 aatgaaaaaaaaaaaaaa 402
            ||||||||||||||||||
Db     1801 AATGAAAAAAAAAAAAAA 1818


RESULT 101
US-08-889-419-1
; Sequence 1, Application US/08889419
```

```
; Patent No. 6071708
;  GENERAL INFORMATION:
;    APPLICANT:  JONES, WILLIAM K
;    APPLICANT:  TUCKER, RONALD F
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    TITLE OF INVENTION:  NOVEL MORPHOGENIC PROTEIN COMPOSITIONS
;    TITLE OF INVENTION:  OF MATTER
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patent Administrator, Testa, Hurwitz &
;      ADDRESSEE:  Thibeault, LLP
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/889,419
;      FILING DATE:  08-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/459,346
;      FILING DATE:  02-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, THOMAS C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  CRP-081DVCN
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
```



```
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
US-08-889-419-1


  Query Match             0.3%;  Score 18;  DB 3;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 102
US-08-445-467-16
; Sequence 16, Application US/08445467
; Patent No. 6077823
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  PANG, ROY HL
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  COHEN, CHARLES M
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  SMART, JOHN E
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED MODULATION OF
;    TITLE OF INVENTION:  INFLAMMATORY RESPONSE
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  35 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/445,467
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/165,511
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/938,336
```



```
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/667,274
;      FILING DATE:  11-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/753,059
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/752,764
;      FILING DATE:  30-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-059CP.APP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
US-08-445-467-16


  Query Match             0.3%;  Score 18;  DB 3;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 103
US-08-480-515A-16
; Sequence 16, Application US/08480515A
```



```
; Patent No. 6090776
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  SMART, JOHN E.
;    TITLE OF INVENTION:  MORPHOGEN TREATMENT OF ORGAN TRANSPLANTS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,515A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-068FWC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /product= "HOP-1"
US-08-480-515A-16


  Query Match             0.3%;  Score 18;  DB 3;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 104
US-08-459-129-1
; Sequence 1, Application US/08459129
; Patent No. 6110482
;  GENERAL INFORMATION:
;     APPLICANT:  KHOURI, ROGER K.
;     APPLICANT:  SAMPATH, KUBER T.
;     APPLICANT:  RUEGER, DAVID C.
;     TITLE OF INVENTION:  MANUFACTURE OF AUTOGENOUS REPLACEMENT
;     TITLE OF INVENTION:  BODY PARTS
;     NUMBER OF SEQUENCES:  3
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  53 STATE STREET
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  U.S.A.
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/459,129
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KELLEY, ROBIN D.
;       REGISTRATION NUMBER:  34,637
;       REFERENCE/DOCKET NUMBER:  CRP-101
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
```



```
;     ORIGINAL SOURCE:
;       ORGANISM:  HOMO SAPIENS
;       TISSUE TYPE:  HIPPOCAMPUS
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..1341
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "OP1"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "OP1"
US-08-459-129-1


  Query Match             0.3%;  Score 18;  DB 3;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 105
US-09-219-391-3
; Sequence 3, Application US/09219391
; Patent No. 6153583
;  GENERAL INFORMATION:
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  OZKAYNAK, ENGIN
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  RUEGER, DAVID C.
;     APPLICANT:  PANG, ROY H.L.
;     APPLICANT:  COHEN, CHARLES M.
;     TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;       STREET:  45 SOUTH STREET
;       CITY:  HOPKINTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  01748
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/219,391
;       FILING DATE:
```



```
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/901,200
;       FILING DATE:  28-JUL-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MEYERS, THOMAS C.
;       REGISTRATION NUMBER:  36,989
;       REFERENCE/DOCKET NUMBER:  CRP-076DV2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (508) 435-9001
;       TELEFAX:  (508) 435-6951
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..1341
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "hOP1-PP"
;       OTHER INFORMATION:  /note= "hOP1 cDNA"
US-09-219-391-3


  Query Match             0.3%;  Score 18;  DB 3;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 106
PCT-US90-07654-3
; Sequence 3, Application PC/TUS9007654
;  GENERAL INFORMATION:
;     APPLICANT:  OZKAYNAK, ENGIN
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  RUEGER, DAVID C
;     TITLE OF INVENTION:  OSTEOGENIC DEVICES
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  53 STATE STREET
;       CITY:  BOSTON
;       STATE:  MA
```



```
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US90/07654
;       FILING DATE:  19901228
;       CLASSIFICATION:  156
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER ESQ, EDMUND R
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP-001P5
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  627/248-7100
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  HOMO SAPIENS
;       TISSUE TYPE:  HIPPOCAMPUS
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..1341
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "OP1"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "OP1"
PCT-US90-07654-3


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 107
```

```
PCT-US91-07635-1
; Sequence 1, Application PC/TUS9107635
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/07635
;      FILING DATE:  19911018
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-056PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
```



```
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US91-07635-1


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT 108
PCT-US92-01968-16
; Sequence 16, Application PC/TUS9201968
;  GENERAL INFORMATION:
;    APPLICANT:  COHEN, CHARLES M
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  PANG, ROY HL
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  RUEGER, DAVID C
;    TITLE OF INVENTION:  PROTEIN-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA HURWITZ & THIBEAULT
;      STREET:  EXCHANGE PLACE  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/01968
;      FILING DATE:  19920311
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 667,274
;      FILING DATE:  11-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 752,764
;      FILING DATE:  30-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-052PC
;    TELECOMMUNICATION INFORMATION:
```



```
;      TELEPHONE:  617/248-7000
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US92-01968-16


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT 109
PCT-US93-05446-1
; Sequence 1, Application PC/TUS9305446
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  PROSTHETIC DEVICES HAVING ENHANCED
;    TITLE OF INVENTION:  OSTEOGENIC PROPERTIES
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Creative BioMolecules, Inc.
;      STREET:  35 South Street
;      CITY:  Hopkinton
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```



```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05446
;      FILING DATE:  19930608
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER  ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  STK-057
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US93-05446-1


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT 110
PCT-US93-07189-1
; Sequence 1, Application PC/TUS9307189
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  NOVEL MORPHOGENIC PROTEIN COMPOSITIONS
;    TITLE OF INVENTION:  OF MATTER
```

```
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR/CREATIVE BIOMOLECULES,
;      ADDRESSEE:  INC.
;      STREET:  35 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07189
;      FILING DATE:  19930729
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-081CP
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US93-07189-1


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```



```
Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 111
PCT-US93-07190-16
; Sequence 16, Application PC/TUS9307190
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN-ENRICHED DIETARY COMPOSITION
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  35 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07190
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-071
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
```



```
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US93-07190-16


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 112
PCT-US93-07231-16
; Sequence 16, Application PC/TUS9307231
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED NERVE REGENERATION AND
;    TITLE OF INVENTION:  REPAIR
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  35 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07231
;      FILING DATE:  19930729
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-070
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```



```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US93-07231-16


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 113
PCT-US93-08742-16
; Sequence 16, Application PC/TUS9308742
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED PERIODONTAL TISSUE REGENERATION
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08742
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KELLEY ESQ, ROBIN D.
;       REGISTRATION NUMBER:  34,637
;       REFERENCE/DOCKET NUMBER:  CRP-067
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7477
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1822 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  HOMO SAPIENS
;       TISSUE TYPE:  HIPPOCAMPUS
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..1341
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "OP1"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "OP1"
PCT-US93-08742-16


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT 114
PCT-US93-08808-16
; Sequence 16, Application PC/TUS9308808
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED LIVER REGENERATION
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  45 SOUTH STREET
```



```
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08808
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY ESQ, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-072
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7477
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US93-08808-16


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
```



```
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT 115
PCT-US93-08885-16
; Sequence 16, Application PC/TUS9308885
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN TREATMENT OF GASTROINTESTINAL
;    TITLE OF INVENTION:  ULCERS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08885
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY ESQ, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-074
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7477
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
```



```
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US93-08885-16


  Query Match             0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1801 AATGAAAAAAAAAAAAAA 1818


RESULT 116
PCT-US93-10520-3
; Sequence 3, Application PC/TUS9310520
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/10520
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/667,274
;      FILING DATE:  11-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/752,764
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/753,059
;      FILING DATE:  30-AUG-1991
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/752,857
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/923,780
;      FILING DATE:  31-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/922,813
;      FILING DATE:  31-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-076PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508)435-9001
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "hOP1-PP"
;      OTHER INFORMATION:  /note= "hOP1 cDNA"
PCT-US93-10520-3


  Query Match              0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 117
PCT-US95-05467-9
; Sequence 9, Application PC/TUS9505467
;  GENERAL INFORMATION:
;    APPLICANT:
```



```
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN-SPECIFIC CELL
;    TITLE OF INVENTION:  SURFACE RECEPTORS AND USES THEREFOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, TESTA, HURWITZ &
;      ADDRESSEE:  THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/05467
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-097PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US95-05467-9


  Query Match              0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
```



```
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 118
PCT-US95-06724-1
; Sequence 1, Application PC/TUS9506724
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MANUFACTURE OF AUTOGENOUS REPLACEMENT
;    TITLE OF INVENTION:  BODY PARTS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, TESTA, HURWITZ &
;      ADDRESSEE:  THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06724
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP101PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HOMO SAPIENS
;      TISSUE TYPE:  HIPPOCAMPUS
;    FEATURE:
```



```
;      NAME/KEY:  CDS
;      LOCATION:  49..1341
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "OP1"
PCT-US95-06724-1


  Query Match              0.3%;  Score 18;  DB 4;  Length 1822;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1801 AATGAAAAAAAAAAAAAA 1818


RESULT 119
US-09-150-133-6
; Sequence 6, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
;  APPLICANT: The Board of Regents of the University of Oklahoma
;  TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
;  TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
;  FILE REFERENCE: 5820.504
;  CURRENT APPLICATION NUMBER: US/09/150,133B
;  CURRENT FILING DATE: 1998-09-09
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 6
;    LENGTH: 1855
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-150-133-6


  Query Match              0.3%;  Score 18;  DB 3;  Length 1855;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3261 acccatttacgctcaagt 3278
          ||||||||||||||||||
Db    673 acccatttacgctcaagt 690


RESULT 120
US-09-150-141-6
; Sequence 6, Application US/09150141B
```

```
; Patent No. 6071732
; GENERAL INFORMATION:
;  APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERAES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 6
;   LENGTH: 1855
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-150-141-6


  Query Match             0.3%;  Score 18;  DB 3;  Length 1855;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3261 acccatttacgctcaagt 3278
          ||||||||||||||||||
Db    673 acccatttacgctcaagt 690


RESULT 121
US-07-841-646-24
; Sequence 24, Application US/07841646
; Patent No. 5266683
; . GENERAL INFORMATION:
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  OZKAYNAK, ENGIN
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  RUEGER, DAVID C.
;     APPLICANT:  PANG, ROY H.L.
;     TITLE OF INVENTION:  OSTEOGENIC DEVICES
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  53 STATE STREET
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  U.S.A.
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
```



```
;       APPLICATION NUMBER:  US/07/841,646
;       FILING DATE:  19920221
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 810,560
;       FILING DATE:  20-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 827,052
;       FILING DATE:  28-JAN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 660,162
;       FILING DATE:  22-FEB-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,988
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,849
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 616,374
;       FILING DATE:  21-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 600,024
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 599,543
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 579,865
;       FILING DATE:  07-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 569,920
;       FILING DATE:  20-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 483,913
;       FILING DATE:  22-FEB-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 422,613
;       FILING DATE:  17-OCT-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 315,342
;       FILING DATE:  23-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 232,630
;       FILING DATE:  15-AUG-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 179,460
;       FILING DATE:  08-APR-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
```



```
;       REFERENCE/DOCKET NUMBER:  CRP-001CP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  24:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1873 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  MURIDAE
;       TISSUE TYPE:  EMBRYO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  104..1393
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "MOP1"
;       OTHER INFORMATION:  /note= "MOP1 (CDNA)"
US-07-841-646-24


  Query Match             0.3%;  Score 18;  DB 1;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 122
US-07-901-703-8
; Sequence 8, Application US/07901703
; Patent No. 5344654
;  GENERAL INFORMATION:
;     APPLICANT:  RUEGER, DAVID C
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  OZAKAYNAK, ENGIN
;     TITLE OF INVENTION:  PROSTHETIC DEVICES HAVING ENHANCED
;     TITLE OF INVENTION:  OSTEOGENIC PROPERTIES
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  EXCHANGE PLACE, 53 STATE STREET
;       CITY:  BOSTON
;       STATE:  MA
```



```
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/901,703
;       FILING DATE:  19920616
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, ESQ, EDMUND R
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  STK-057
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1873 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  MURIDAE
;       TISSUE TYPE:  EMBRYO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  104..1393
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "MOP1"
;       OTHER INFORMATION:  /note= "MOP1 (CDNA)"
US-07-901-703-8


  Query Match             0.3%;  Score 18;  DB 1;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 123
US-08-147-023-24
; Sequence 24, Application US/08147023
; Patent No. 5468845
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,023
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
```



```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1"
;      OTHER INFORMATION:  /note= "MOP1 CDNA"
US-08-147-023-24


  Query Match                 0.3%;  Score 18;  DB 1;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 124
```



```
US-08-278-729A-18
; Sequence 18, Application US/08278729A
; Patent No. 5650276
;  GENERAL INFORMATION:
;    APPLICANT:  SMART, JOHN
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN SCREENING METHOD
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/278,729A
;      FILING DATE:  20-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER Esq., EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-058CPFW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /product= "MOP1 (CDNA)"
US-08-278-729A-18


  Query Match                 0.3%;  Score 18;  DB 1;  Length 1873;
```



```
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 125
US-08-480-528A-5
; Sequence 5, Application US/08480528A
; Patent No. 5652118
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,528A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-076FW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "mOP1-PP"
; OTHER INFORMATION: /note= "mOP1 cDNA"
US-08-480-528A-5

Query Match 0.3%; Score 18; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 126
US-08-479-666-5
; Sequence 5, Application US/08479666
; Patent No. 5652337
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; APPLICANT: COHEN, CHARLES M.
; TITLE OF INVENTION: OP1-INDUCED MORPHOGENESIS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,666
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FENTON Esq., GILLIAN M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: CRP-076DV

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 435-9001
; TELEFAX: (508) 435-6951
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "mOP1-PP"
; OTHER INFORMATION: /note= "mOP1 cDNA"
US-08-479-666-5

Query Match 0.3%; Score 18; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 127
US-08-155-343A-18
; Sequence 18, Application US/08155343A
; Patent No. 5656593
; GENERAL INFORMATION:
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: OPPERMANN, HERMAN
; APPLICANT: COHEN, CHARLES M.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: MORPHOGENIC-INDUCED PERIODONTAL TISSUE
; TITLE OF INVENTION: REGENERATION.
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/155,343A
; FILING DATE: 15-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FENTON Esq., GILLIAN M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: CRP-067FW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7560
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /product= "MOP1 (CDNA)"
US-08-155-343A-18

Query Match 0.3%; Score 18; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 128
US-08-406-672-18
; Sequence 18, Application US/08406672
; Patent No. 5674844
; GENERAL INFORMATION:
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: COHEN, CHARLES M.
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: TREATMENT TO PREVENT LOSS OF AND/OR
; TITLE OF INVENTION: INCREASE BONE MASS IN METABOLIC BONE DISEASES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/406,672
; FILING DATE: 20-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 752,857
; FILING DATE: 30-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 667,274
; FILING DATE: 11-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: FENTON Esq., GILLIAN M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: CRP-060CN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7560
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /product= "MOP1 (CDNA)"
US-08-406-672-18

Query Match 0.3%; Score 18; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

```
RESULT 129
US-08-643-563A-18
; Sequence 18, Application US/08643563A
; Patent No. 5707810
;  GENERAL INFORMATION:
;    APPLICANT:  SMART, JOHN
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGENIC PROTEIN SCREENING METHOD
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/643,563A
;      FILING DATE:  06-MAY-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  TWOMEY Esq., MICHAEL J.
;      REGISTRATION NUMBER:  38,349
;      REFERENCE/DOCKET NUMBER:  CRP-058CN2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /product= "MOP1 (CDNA)"
US-08-643-563A-18
```



```
  Query Match           0.3%;  Score 18;  DB 1;  Length 1873;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 130
US-08-447-570-24
; Sequence 24, Application US/08447570
; Patent No. 5714589
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,570
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
```



```
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1"
;      OTHER INFORMATION:  /note= "MOP1 CDNA"
```



```
US-08-447-570-24


  Query Match           0.3%;  Score 18;  DB 1;  Length 1873;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 131
US-08-643-763A-18
; Sequence 18, Application US/08643763A
; Patent No. 5733878
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  MORPHOGENIC-INDUCED PERIODONTAL TISSUE
;    TITLE OF INVENTION:  REGENERATION.
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/643,763A
;      FILING DATE:  06-MAY-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-067CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /product= "MOP1 (CDNA)"
US-08-643-763A-18

Query Match 0.3%; Score 18; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 132
US-08-462-623-18
; Sequence 18, Application US/08462623
; Patent No. 5739107
; GENERAL INFORMATION:
; APPLICANT: COHEN, CHARLES M.
; APPLICANT: CHARETTE, MARC F.
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: PANG, ROY H.L.
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: SMART, JOHN E.
; TITLE OF INVENTION: MORPHOGEN TREATMENT OF GASTROINTESTINAL
; TITLE OF INVENTION: ULCERS.
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,623
; FILING DATE:

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/445,882
; FILING DATE: 22-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FENTON Esq., GILLIAN M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: CRP-074CN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 435-9001
; TELEFAX: (508) 435-6951
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /product= "MOP1 (CDNA)"
US-08-462-623-18

Query Match 0.3%; Score 18; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 133
US-08-451-953A-18
; Sequence 18, Application US/08451953A
; Patent No. 5741641
; GENERAL INFORMATION:
; APPLICANT: SMART, JOHN
; APPLICANT: OPPERMANN, HERMAN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; APPLICANT: COHEN, CHARLES M.
; TITLE OF INVENTION: MORPHOGENIC PROTEIN SCREENING METHOD
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET

; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,953A
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER Esq., EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-058CN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 435-9001
; TELEFAX: (508) 435-6951
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /product= "MOP1 (CDNA)"
US-08-451-953A-18

Query Match 0.3%; Score 18; DB 1; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 134
US-08-459-346-3
; Sequence 3, Application US/08459346
; Patent No. 5834179
; GENERAL INFORMATION:
; APPLICANT: JONES, WILLIAM K
; APPLICANT: TUCKER, RONALD F
; APPLICANT: RUEGER, DAVID C

; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; TITLE OF INVENTION: NOVEL MORPHOGENIC PROTEIN COMPOSITIONS
; TITLE OF INVENTION: OF MATTER
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR/CREATIVE BIOMOLECULES,
; ADDRESSEE: INC.
; STREET: 35 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,346
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/029,335
; FILING DATE: 04-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,091
; FILING DATE: 03-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/946,235
; FILING DATE: 16-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,336
; FILING DATE: 08-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/923,780
; FILING DATE: 31-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-081CP
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS

```
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1"
;      OTHER INFORMATION:  /note= "MOP1 cDNA"
US-08-459-346-3


  Query Match              0.3%;  Score 18;  DB 2;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 135
US-08-445-468A-18
; Sequence 18, Application US/08445468A
; Patent No. 5849686
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED LIVER REGENERATION
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/445,468A
;      FILING DATE:  22-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  CRP-072FW2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
```



```
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /product= "MOP1 (CDNA)"
US-08-445-468A-18


  Query Match              0.3%;  Score 18;  DB 2;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 136
US-08-901-200A-5
; Sequence 5, Application US/08901200A
; Patent No. 5854071
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    APPLICANT:  COHEN, CHARLES M.
;    TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/901,200A
;      FILING DATE:  28-JUL-1997
```



```
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, THOMAS C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  CRP-076DV2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 435-9001
;      TELEFAX:  (508) 435-6951
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "mOP1-PP"
;      OTHER INFORMATION:  /note= "mOP1 cDNA"
US-08-901-200A-5


  Query Match              0.3%;  Score 18;  DB 2;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 137
US-08-449-700-24
; Sequence 24, Application US/08449700
; Patent No. 5863758
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
```



```
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/449,700
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
```

; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "MOP1"
; OTHER INFORMATION: /note= "MOP1 CDNA"
US-08-449-700-24

Query Match 0.3%; Score 18; DB 2; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 138
US-08-449-699A-24
; Sequence 24, Application US/08449699A
; Patent No. 5958441
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: ANTIBODIES TO OSTEOGENIC PROTEINS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,699A
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/147,023
; FILING DATE: 01-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: STK-001CP6CN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "MOP1"
; OTHER INFORMATION: /note= "MOP1 CDNA"
US-08-449-699A-24

Query Match 0.3%; Score 18; DB 2; Length 1873;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 139
US-08-461-397A-18
; Sequence 18, Application US/08461397A

; Patent No. 5972884
; GENERAL INFORMATION:
; APPLICANT: COHEN, CHARLES M.
; APPLICANT: CHARETTE, MARC F.
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: PANG, ROY H.L.
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: SMART, JOHN E.
; TITLE OF INVENTION: MORPHOGEN TREATMENT FOR LIMITING
; TITLE OF INVENTION: PROLIFERATION OF EPITHELIAL CELLS.
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,397A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-074FW2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1393
; OTHER INFORMATION: /product= "MOP1 (CDNA)"
US-08-461-397A-18

Query Match 0.3%; Score 18; DB 2; Length 1873;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1835 AATGAAAAAAAAAAAAAA 1852

RESULT 140
US-08-912-088-18
; Sequence 18, Application US/08912088
; Patent No. 5994131
; GENERAL INFORMATION:
; APPLICANT: SMART, JOHN
; APPLICANT: OPPERMANN, HERMAN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; APPLICANT: COHEN, CHARLES M.
; TITLE OF INVENTION: MORPHOGENIC PROTEIN SCREENING METHOD
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/912,088
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,729
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER Esq., EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-058CPFW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 435-9001
; TELEFAX: (508) 435-6951
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1873 base pairs

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  104..1393
;       OTHER INFORMATION:  /product= "MOP1 (CDNA)"
US-08-912-088-18


  Query Match             0.3%;  Score 18;  DB 2;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 141
US-08-278-730A-18
; Sequence 18, Application US/08278730A
; Patent No. 6022853
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  COHEN, CHARLES M.
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  OPPERMANN, HERMAN
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  MORPHOGENIC-ENRICHED DIETARY COMPOSITION
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/278,730A
;      FILING DATE:  20-JULY-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FENTON Esq., GILLIAN M.
;      REGISTRATION NUMBER:  36,508
```



```
;      REFERENCE/DOCKET NUMBER:  CRP-071PW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7560
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /product= "MOP1 (CDNA)"
US-08-278-730A-18


  Query Match             0.3%;  Score 18;  DB 3;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 142
US-08-889-419-3
; Sequence 3, Application US/08889419
; Patent No. 6071708
;  GENERAL INFORMATION:
;    APPLICANT:  JONES, WILLIAM K
;    APPLICANT:  TUCKER, RONALD F
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    TITLE OF INVENTION:  NOVEL MORPHOGENIC PROTEIN COMPOSITIONS
;    TITLE OF INVENTION:  OF MATTER
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patent Administrator, Testa, Hurwitz &
;      ADDRESSEE:  Thibeault, LLP
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```



```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/889,419
;       FILING DATE:  08-JUL-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/459,346
;       FILING DATE:  02-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MEYERS, THOMAS C.
;       REGISTRATION NUMBER:  36,989
;       REFERENCE/DOCKET NUMBER:  CRP-081DVCN
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1873 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  104..1393
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "MOP1"
;       OTHER INFORMATION:  /note= "MOP1 CDNA"
US-08-889-419-3


  Query Match             0.3%;  Score 18;  DB 3;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    1835 AATGAAAAAAAAAAAAAA 1852


RESULT 143
US-08-445-467-18
; Sequence 18, Application US/08445467
; Patent No. 6077823
;  GENERAL INFORMATION:
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  PANG, ROY HL
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  RUEGER, DAVID C
;    APPLICANT:  COHEN, CHARLES M
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  SMART, JOHN E
```



```
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED MODULATION OF
;    TITLE OF INVENTION:  INFLAMMATORY RESPONSE
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  35 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/445,467
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/165,511
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/938,336
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/667,274
;      FILING DATE:  11-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/753,059
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/752,764
;      FILING DATE:  30-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-059CP.APP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MURIDAE
;      TISSUE TYPE:  EMBRYO
```

```
;       NAME/KEY:  CDS
;       LOCATION:  104..1393
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "MOP1"
;       OTHER INFORMATION:  /note= "MOP1 (CDNA)"
US-08-445-467-18

  Query Match                 0.3%;  Score 18;  DB 3;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 144
US-08-480-515A-18
; Sequence 18, Application US/08480515A
; Patent No. 6090776
; GENERAL INFORMATION:
;   APPLICANT:  KUBERASAMPATH, THANGAVEL
;   APPLICANT:  PANG, ROY H.L.
;   APPLICANT:  OPPERMANN, HERMANN
;   APPLICANT:  RUEGER, DAVID C.
;   APPLICANT:  COHEN, CHARLES M.
;   APPLICANT:  SMART, JOHN E.
;   TITLE OF INVENTION:  MORPHOGEN TREATMENT OF ORGAN TRANSPLANTS
;   NUMBER OF SEQUENCES:  33
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;     STREET:  45 SOUTH STREET
;     CITY:  HOPKINTON
;     STATE:  MA
;     COUNTRY:  USA
;     ZIP:  01748
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/480,515A
;     FILING DATE:  07-JUN-1995
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  FENTON Esq., GILLIAN M.
;     REGISTRATION NUMBER:  36,508
;     REFERENCE/DOCKET NUMBER:  CRP-068FWC
```



```
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/219,391
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/901,200
;     FILING DATE:  28-JUL-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  MEYERS, THOMAS C.
;     REGISTRATION NUMBER:  36,989
;     REFERENCE/DOCKET NUMBER:  CRP-076DV2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (508) 435-9001
;     TELEFAX:  (508) 435-6951
; INFORMATION FOR SEQ ID NO:  5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1873 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  104..1393
;     OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;     OTHER INFORMATION:  /product= "mOP1-PP"
;     OTHER INFORMATION:  /note= "mOP1 cDNA"
US-09-219-391-5

  Query Match                 0.3%;  Score 18;  DB 3;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 146
PCT-US90-07654-1
; Sequence 1, Application PC/TUS9007654
; GENERAL INFORMATION:
;   APPLICANT:  OZKAYNAK, ENGIN
;   APPLICANT:  OPPERMANN, HERMANN
;   APPLICANT:  KUBERASAMPATH, THANGAVEL
;   APPLICANT:  RUEGER, DAVID C
;   TITLE OF INVENTION:  OSTEOGENIC DEVICES
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
```



```
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (617) 248-7560
;     TELEFAX:  (617) 248-7100
; INFORMATION FOR SEQ ID NO:  18:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1873 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  104..1393
;     OTHER INFORMATION:  /product= "MOP1 (CDNA)"
US-08-480-515A-18

  Query Match                 0.3%;  Score 18;  DB 3;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 145
US-09-219-391-5
; Sequence 5, Application US/09219391
; Patent No. 6153583
; GENERAL INFORMATION:
;   APPLICANT:  OPPERMANN, HERMANN
;   APPLICANT:  OZKAYNAK, ENGIN
;   APPLICANT:  KUBERASAMPATH, THANGAVEL
;   APPLICANT:  RUEGER, DAVID C.
;   APPLICANT:  PANG, ROY H.L.
;   APPLICANT:  COHEN, CHARLES M.
;   TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;   NUMBER OF SEQUENCES:  15
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;     STREET:  45 SOUTH STREET
;     CITY:  HOPKINTON
;     STATE:  MA
;     COUNTRY:  USA
;     ZIP:  01748
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
```



```
;     STREET:  53 STATE STREET
;     CITY:  BOSTON
;     STATE:  MA
;     COUNTRY:  USA
;     ZIP:  02109
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/US90/07654
;     FILING DATE:  19901228
;     CLASSIFICATION:  156
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  PITCHER ESQ, EDMUND R
;     REGISTRATION NUMBER:  27,829
;     REFERENCE/DOCKET NUMBER:  CRP-001P5
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  617/248-7000
;     TELEFAX:  627/248-7100
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1873 base pairs
;     TYPE:  NUCLEIC ACID
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   ORIGINAL SOURCE:
;     ORGANISM:  NURIDAE
;     TISSUE TYPE:  EMBRYO
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  104..1393
;     OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;     OTHER INFORMATION:  /product= "MOP1"
;     OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US90-07654-1

  Query Match                 0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852
```

```
RESULT 147
PCT-US92-01968-18
; Sequence 18, Application PC/TUS9201968
;  GENERAL INFORMATION:
;     APPLICANT:  COHEN, CHARLES M
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  PANG, ROY HL
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  RUEGER, DAVID C
;     TITLE OF INVENTION:  PROTEIN-INDUCED MORPHOGENESIS
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA HURWITZ & THIBEAULT
;       STREET:  EXCHANGE PLACE  53 STATE STREET
;       CITY:  BOSTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US92/01968
;       FILING DATE:  19920311
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 667,274
;       FILING DATE:  11-MAR-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 752,764
;       FILING DATE:  30-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER ESQ, EDMUND R
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP-052PC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1873 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  MURIDAE
;       TISSUE TYPE:  EMBRYO
```



```
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  104..1393
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "MOP1"
;       OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US92-01968-18


  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 148
PCT-US93-05446-8
; Sequence 8, Application PC/TUS9305446
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  PROSTHETIC DEVICES HAVING ENHANCED
;     TITLE OF INVENTION:  OSTEOGENIC PROPERTIES
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Creative BioMolecules, Inc.
;       STREET:  35 South Street
;       CITY:  Hopkinton
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  01748
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/05446
;       FILING DATE:  19930608
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER  ESQ, EDMUND R
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  STK-057
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1873 base pairs
```



```
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  MURIDAE
;       TISSUE TYPE:  EMBRYO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  104..1393
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "MOP1"
;       OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US93-05446-8


  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 149
PCT-US93-07189-3
; Sequence 3, Application PC/TUS9307189
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  NOVEL MORPHOGENIC PROTEIN COMPOSITIONS
;     TITLE OF INVENTION:  OF MATTER
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  PATENT ADMINISTRATOR/CREATIVE BIOMOLECULES,
;       ADDRESSEE:  INC.
;       STREET:  35 SOUTH STREET
;       CITY:  HOPKINTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  01748
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/07189
;       FILING DATE:  19930729
```



```
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KELLEY, ROBIN D.
;       REGISTRATION NUMBER:  34,637
;       REFERENCE/DOCKET NUMBER:  CRP-081CP
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1873 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  104..1393
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "MOP1"
;       OTHER INFORMATION:  /note= "MOP1 CDNA"
PCT-US93-07189-3


  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 150
PCT-US93-07190-18
; Sequence 18, Application PC/TUS9307190
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  MORPHOGEN-ENRICHED DIETARY COMPOSITION
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;       STREET:  35 SOUTH STREET
;       CITY:  HOPKINTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  01748
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07190
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-071
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MURIDAE
;      TISSUE TYPE:  EMBRYO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1"
;      OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US93-07190-18


  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 151
PCT-US93-07231-18
; Sequence 18, Application PC/TUS9307231
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED NERVE REGENERATION AND
;    TITLE OF INVENTION:  REPAIR
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  35 SOUTH STREET
```



```
;      CITY:  HOPKINTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07231
;      FILING DATE:  19930729
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-070
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MURIDAE
;      TISSUE TYPE:  EMBRYO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1"
;      OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US93-07231-18


  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 152
```



```
PCT-US93-08742-18
; Sequence 18, Application PC/TUS9308742
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED PERIODONTAL TISSUE REGENERATION
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08742
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY ESQ, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-067
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7477
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MURIDAE
;      TISSUE TYPE:  EMBRYO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1"
;      OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US93-08742-18
```



```
  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 153
PCT-US93-08808-18
; Sequence 18, Application PC/TUS9308808
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN-INDUCED LIVER REGENERATION
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08808
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELLEY ESQ, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-072
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7477
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
```

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MURIDAE
;      TISSUE TYPE:  EMBRYO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1"
;      OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US93-08608-18


  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 154
PCT-US93-08885-18
; Sequence 18, Application PC/TUS9308885
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MORPHOGEN TREATMENT OF GASTROINTESTINAL
;    TITLE OF INVENTION:  ULCERS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08885
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
```



```
;      NAME:  KELLEY ESQ, ROBIN D.
;      REGISTRATION NUMBER:  34,637
;      REFERENCE/DOCKET NUMBER:  CRP-074.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7477
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MURIDAE
;      TISSUE TYPE:  EMBRYO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1"
;      OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US93-08885-18


  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 155
PCT-US93-10520-5
; Sequence 5, Application PC/TUS9310520
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  OP3-INDUCED MORPHOGENESIS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CREATIVE BIOMOLECULES, INC.
;      STREET:  45 SOUTH STREET
;      CITY:  HOPKINTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  01748
;    COMPUTER READABLE FORM:
```



```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/10520
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/667,274
;      FILING DATE:  11-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/752,764
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/753,059
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/752,857
;      FILING DATE:  30-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/923,780
;      FILING DATE:  31-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/922,813
;      FILING DATE:  31-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-076PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508)435-9001
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1873 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MURIDAE
;      TISSUE TYPE:  EMBRYO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  104..1393
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "MOP1-PP"
;      OTHER INFORMATION:  /note= "MOP1 (CDNA)"
PCT-US93-10520-5
```



```
  Query Match             0.3%;  Score 18;  DB 4;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   1835 AATGAAAAAAAAAAAAAA 1852


RESULT 156
US-08-458-120-1
; Sequence 1, Application US/08458120
; Patent No. 5635616
;  GENERAL INFORMATION:
;    APPLICANT:  OLSEN, ET AL.
;    TITLE OF INVENTION:  Human Oxalyl-CoA Decarboxylase
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,120
;      FILING DATE:  SUBMITTED HEREWITH
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/05561
;      FILING DATE:  18 MAY 94
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FERRARO, GREGORY D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-312
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1882 BASE PAIRS
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
```

; TOPOLOGY: LINEAR
; MOLECULE TYPE: cDNA
US-08-458-120-1

Query Match 0.3%; Score 18; DB 1; Length 1882;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1861 AATGAAAAAAAAAAAAAA 1878

RESULT 157
US-08-867-970-1
; Sequence 1, Application US/08867970
; Patent No. 5945273
; GENERAL INFORMATION:
; APPLICANT: Olsen, Henrik S.
; APPLICANT: Coleman, Timothy
; APPLICANT: Adams, Mark D.
; TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,970
; FILING DATE: JUN-3-97
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/458,120
; FILING DATE: 02-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/05561
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Robert H.
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PF119D1
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10..1744
US-08-867-970-1

Query Match 0.3%; Score 18; DB 2; Length 1882;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1861 AATGAAAAAAAAAAAAAA 1878

RESULT 158
US-08-387-117-7
; Sequence 7, Application US/08387117
; Patent No. 6127165
; GENERAL INFORMATION:
; APPLICANT: Huang, Chin-shiou
; APPLICANT: Meister, Alton
; APPLICANT: Anderson, Mary E.
; TITLE OF INVENTION: Rat Glutathione Synthetase Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,117
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:

; NAME: Dolak, Lisa A.
; REGISTRATION NUMBER: 35,491
; REFERENCE/DOCKET NUMBER: 19603/610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1630
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1905 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-387-117-7

Query Match 0.3%; Score 18; DB 3; Length 1905;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1879 AATGAAAAAAAAAAAAAA 1896

RESULT 159
US-08-900-148-1
; Sequence 1, Application US/08900148
; Patent No. 5965792
; GENERAL INFORMATION:
; APPLICANT: Schroeder, Julian I.
; APPLICANT: Antosiewicz, Danuta M.
; APPLICANT: Schachtman, Daniel P.
; APPLICANT: Clemens, Stephan
; TITLE OF INVENTION: Nucleic Acids Encoding Metal Uptake
; TITLE OF INVENTION: Transporters and Their Uses
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,148

; FILING DATE: 28-JUL-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/022,722
; FILING DATE: 29-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-087600US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1982 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 137..1471
; OTHER INFORMATION: /product= "low affinity cation
; OTHER INFORMATION: transporter 1 (LCT1)"
US-08-900-148-1

Query Match 0.3%; Score 18; DB 2; Length 1982;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 1956 AATGAAAAAAAAAAAAAA 1973

RESULT 160
US-08-096-277-6
; Sequence 6, Application US/08096277
; Patent No. 5578482
; GENERAL INFORMATION:
; APPLICANT: Lippman, Marc E
; APPLICANT: Lupu, Ruth
; TITLE OF INVENTION: Ligand Growth Factors that Bind to the
; TITLE OF INVENTION: erbB-2 Receptor Protein and Induce Cellular Response
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.

```
;       COUNTRY:  U.S.
;       ZIP:  20001
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/096,277
;       FILING DATE:  26-JUL-1993
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/875,788
;       FILING DATE:  29-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/640,497
;       FILING DATE:  14-JAN-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/917,988
;       FILING DATE:  24-JUL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/872,114
;       FILING DATE:  22-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/528,438
;       FILING DATE:  25-MAY-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hoscheit, Dale H
;       REGISTRATION NUMBER:  19,090
;       REFERENCE/DOCKET NUMBER:  02899.43360
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-508-9100
;       TELEFAX:  202-508-9299
;       TELEX:  197430 BBMB UT
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2164 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
US-08-096-277-6


  Query Match             0.3%;  Score 18;  DB 1;  Length 2164;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2145 AATGAAAAAAAAAAAAAA 2162
```



```
RESULT 161
US-08-550-815-6
; Sequence 6, Application US/08550815
; Patent No. 5869618
;  GENERAL INFORMATION:
;    APPLICANT:  Lippman, Marc E
;    APPLICANT:  Lupu, Ruth
;    TITLE OF INVENTION:  Ligand Growth Factors that Bind to the
;    TITLE OF INVENTION:  erbB-2 Receptor Protein and Induce Cellular Response
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckeet
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/550,815
;      FILING DATE:  31-OCT-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/096,277
;      FILING DATE:  26-JUL-1993
;      APPLICATION NUMBER:  US 07/875,788
;      FILING DATE:  29-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/640,497
;      FILING DATE:  14-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/917,988
;      FILING DATE:  24-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,114
;      FILING DATE:  22-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/528,438
;      FILING DATE:  25-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hoscheit, Dale H
;      REGISTRATION NUMBER:  19,090
;      REFERENCE/DOCKET NUMBER:  02899.43360
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
```



```
;       TELEFAX:  202-508-9299
;       TELEX:  197430 BBMB UT
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2164 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
US-08-550-815-6


  Query Match             0.3%;  Score 18;  DB 2;  Length 2164;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2145 AATGAAAAAAAAAAAAAA 2162


RESULT 162
US-08-703-089-6
; Sequence 6, Application US/08703089
; Patent No. 6040290
;  GENERAL INFORMATION:
;    APPLICANT:  Lippman, Marc E
;    APPLICANT:  Lupu, Ruth
;    TITLE OF INVENTION:  Ligand Growth Factors that Bind to the
;    TITLE OF INVENTION:  erbB-2 Receptor Protein and Induce Cellular Response
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckeet
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/703,089
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/096,277
;      FILING DATE:  26-JUL-1993
;      APPLICATION NUMBER:  US 07/875,788
```



```
;       FILING DATE:  29-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/640,497
;       FILING DATE:  14-JAN-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/917,988
;       FILING DATE:  24-JUL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/872,114
;       FILING DATE:  22-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/528,438
;       FILING DATE:  25-MAY-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hoscheit, Dale H
;       REGISTRATION NUMBER:  19,090
;       REFERENCE/DOCKET NUMBER:  02899.43360
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-508-9100
;       TELEFAX:  202-508-9299
;       TELEX:  197430 BBMB UT
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2164 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
US-08-703-089-6


  Query Match             0.3%;  Score 18;  DB 3;  Length 2164;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2145 AATGAAAAAAAAAAAAAA 2162


RESULT 163
US-07-847-743B-7
; Sequence 7, Application US/07847743B
; Patent No. 5367060
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    TITLE OF INVENTION:  Structure, Production and Use of
;    TITLE OF INVENTION:  Heregulin
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
```

; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/847,743B
; FILING DATE: 19920306
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705256
; FILING DATE: 24-May-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/765212
; FILING DATE: 25-Sep-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/790801
; FILING DATE: 08-No. 5367060-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hensley, Max D.
; REGISTRATION NUMBER: 27,043
; REFERENCE/DOCKET NUMBER: 712P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1489
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-847-743B-7

Query Match 0.3%; Score 18; DB 1; Length 2199;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 2180 AATGAAAAAAAAAAAAAA 2197

RESULT 164
US-08-096-277-7

; Sequence 7, Application US/08096277
; Patent No. 5578482
; GENERAL INFORMATION:
; APPLICANT: Lippman, Marc E
; APPLICANT: Lupu, Ruth
; TITLE OF INVENTION: Ligand Growth Factors that Bind to the
; TITLE OF INVENTION: erbB-2 Receptor Protein and Induce Cellular Response
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,277
; FILING DATE: 26-JUL-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/875,788
; FILING DATE: 29-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/640,497
; FILING DATE: 14-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,988
; FILING DATE: 24-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,114
; FILING DATE: 22-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/528,438
; FILING DATE: 25-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoscheit, Dale H
; REGISTRATION NUMBER: 19,090
; REFERENCE/DOCKET NUMBER: 02899.43360
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-096-277-7

Query Match 0.3%; Score 18; DB 1; Length 2199;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 2180 AATGAAAAAAAAAAAAAA 2197

RESULT 165
US-08-456-201-7
; Sequence 7, Application US/08456201
; Patent No. 5641869
; GENERAL INFORMATION:
; APPLICANT: Vandlen, Richard L.,
; APPLICANT: Holmes, William E.
; TITLE OF INVENTION: Structure, Production and Use of
; TITLE OF INVENTION: Heregulin 2 Ligands
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,201
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/126,145
; FILING DATE:
; APPLICATION NUMBER: 07/880,917
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705256
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/765212
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/790801

; ATTORNEY/AGENT INFORMATION:
; NAME: Hensley, Max D.
; REGISTRATION NUMBER: 27,034
; REFERENCE/DOCKET NUMBER: 712P4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1489
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-456-201-7

Query Match 0.3%; Score 18; DB 1; Length 2199;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 aatgaaaaaaaaaaaaaa 402
||||||||||||||||||
Db 2180 AATGAAAAAAAAAAAAAA 2197

RESULT 166
US-08-456-241-7
; Sequence 7, Application US/08456241
; Patent No. 5840525
; GENERAL INFORMATION:
; APPLICANT: Vandlen, Richard L.
; APPLICANT: Holmes, William E.
; TITLE OF INVENTION: STRUCTURE, PRODUCTION AND USE OF
; TITLE OF INVENTION: HEREGULIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,241
; FILING DATE: 31-MAY-1995

```
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/126145
;       FILING DATE:  23-SEP-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/880917
;       FILING DATE:  11-MAY-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/847743
;       FILING DATE:  06-MAR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/790801
;       FILING DATE:  08-NOV-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/765212
;       FILING DATE:  25-SEP-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/705256
;       FILING DATE:  24-MAY-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lee, Wendy M.
;       REGISTRATION NUMBER:  00,000
;       REFERENCE/DOCKET NUMBER:  712P4C1D1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-1994
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2199 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-456-241-7


  Query Match             0.3%;  Score 18;  DB 2;  Length 2199;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2180 AATGAAAAAAAAAAAAAA 2197


RESULT 167
US-08-550-815-7
; Sequence 7, Application US/08550815
; Patent No. 5869618
;  GENERAL INFORMATION:
;     APPLICANT:  Lippman, Marc E
```



```
;     APPLICANT:  Lupu, Ruth
;     TITLE OF INVENTION:  Ligand Growth Factors that Bind to the
;     TITLE OF INVENTION:  erbB-2 Receptor Protein and Induce Cellular Response
;     NUMBER OF SEQUENCES:  18
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Banner, Birch, McKie & Beckeet
;       STREET:  1001 G Street, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.
;       ZIP:  20001
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/550,815
;       FILING DATE:  31-OCT-1995
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/096,277
;       FILING DATE:  26-JUL-1993
;       APPLICATION NUMBER:  US 07/875,788
;       FILING DATE:  29-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/640,497
;       FILING DATE:  14-JAN-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/917,988
;       FILING DATE:  24-JUL-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/872,114
;       FILING DATE:  22-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/528,438
;       FILING DATE:  25-MAY-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hoscheit, Dale H
;       REGISTRATION NUMBER:  19,090
;       REFERENCE/DOCKET NUMBER:  02899.43360
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-508-9100
;       TELEFAX:  202-508-9299
;       TELEX:  197430 BBMB UT
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2199 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  unknown
```



```
;     MOLECULE TYPE:  DNA (genomic)
US-08-550-815-7


  Query Match             0.3%;  Score 18;  DB 2;  Length 2199;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2180 AATGAAAAAAAAAAAAAA 2197


RESULT 168
US-08-703-089-7
; Sequence 7, Application US/08703089
; Patent No. 6040290
;  GENERAL INFORMATION:
;     APPLICANT:  Lippman, Marc E
;     APPLICANT:  Lupu, Ruth
;     TITLE OF INVENTION:  Ligand Growth Factors that Bind to the
;     TITLE OF INVENTION:  erbB-2 Receptor Protein and Induce Cellular Response
;     NUMBER OF SEQUENCES:  18
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Banner, Birch, McKie & Beckeet
;       STREET:  1001 G Street, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.
;       ZIP:  20001
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/703,089
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/096,277
;       FILING DATE:  26-JUL-1993
;       APPLICATION NUMBER:  US 07/875,788
;       FILING DATE:  29-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/640,497
;       FILING DATE:  14-JAN-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/917,988
;       FILING DATE:  24-JUL-1992
;     PRIOR APPLICATION DATA:
```



```
;       APPLICATION NUMBER:  US 07/872,114
;       FILING DATE:  22-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/528,438
;       FILING DATE:  25-MAY-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hoscheit, Dale H
;       REGISTRATION NUMBER:  19,090
;       REFERENCE/DOCKET NUMBER:  02899.43360
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-508-9100
;       TELEFAX:  202-508-9299
;       TELEX:  197430 BBMB UT
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2199 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-08-703-089-7


  Query Match             0.3%;  Score 18;  DB 3;  Length 2199;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2180 AATGAAAAAAAAAAAAAA 2197


RESULT 169
PCT-US92-04295A-7
; Sequence 7, Application PC/TUS9204295A
;  GENERAL INFORMATION:
;     APPLICANT:  Genentech, Inc.
;     TITLE OF INVENTION:  Structure, Production and Use of
;     TITLE OF INVENTION:  Heregulin
;     NUMBER OF SEQUENCES:  30
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genentech, Inc.
;       STREET:  460 Point San Bruno Blvd
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  patin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US92/04295A
;       FILING DATE:  19920521
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/705256
;       FILING DATE:  24-MAY-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/765212
;       FILING DATE:  25-SEP-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/790801
;       FILING DATE:  08-NOV-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/847743
;       FILING DATE:  06-MAR-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hensley, Max D.
;       REGISTRATION NUMBER:  27,043
;       REFERENCE/DOCKET NUMBER:  712P4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-1994
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2199 bases
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
PCT-US92-04295A-7


  Query Match           0.3%;  Score 18;  DB 4;  Length 2199;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2180 AATGAAAAAAAAAAAAAA 2197


RESULT 170
US-08-317-522A-1
; Sequence 1, Application US/08317522A
; Patent No. 5599918
;  GENERAL INFORMATION:
;     APPLICANT:  Fukuda, Michiko N.
;     TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;     TITLE OF INVENTION:  Proteins
```



```
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Campbell and Flores
;       STREET:  4370 La Jolla Village Drive, Suite 700
;       CITY:  San Diego
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  92122
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/317,522A
;       FILING DATE:  04-OCT-1994
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryn A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-LA 9991
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2524 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  28..2275
US-08-317-522A-1


  Query Match           0.3%;  Score 18;  DB 1;  Length 2524;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2486 AATGAAAAAAAAAAAAAA 2503


RESULT 171
US-08-439-818A-1
; Sequence 1, Application US/08439818A
; Patent No. 5654145
;  GENERAL INFORMATION:
;     APPLICANT:  Fukuda, Michiko N.
```



```
;     TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;     TITLE OF INVENTION:  Proteins
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Campbell and Flores
;       STREET:  4370 La Jolla Village Drive, Suite 700
;       CITY:  San Diego
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  92122
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/439,818A
;       FILING DATE:  12-MAY-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/317,522
;       FILING DATE:  04-OCT-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryn A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-LA 1563
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2524 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  28..2275
US-08-439-818A-1


  Query Match           0.3%;  Score 18;  DB 1;  Length 2524;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2486 AATGAAAAAAAAAAAAAA 2503


RESULT 172
```



```
US-08-751-965-1
; Sequence 1, Application US/08751965
; Patent No. 5858360
;  GENERAL INFORMATION:
;     APPLICANT:  Fukuda, Michiko N.
;     TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;     TITLE OF INVENTION:  Proteins
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Campbell and Flores
;       STREET:  4370 La Jolla Village Drive, Suite 700
;       CITY:  San Diego
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  92122
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/751,965
;       FILING DATE:  Herewith
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/439,818
;       FILING DATE:  12-MAY-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryn A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-LA 2252
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2524 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  28..2275
US-08-751-965-1


  Query Match           0.3%;  Score 18;  DB 2;  Length 2524;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
```

```
                   ||||||||||||||||||
Db    2486 AATGAAAAAAAAAAAAAA 2503


RESULT 173
US-08-738-975-1
; Sequence 1, Application US/08738975
; Patent No. 5880267
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/738,975
;      FILING DATE:  herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/439,818
;      FILING DATE:  05-Dec-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 2251
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2524 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..2275
US-08-738-975-1
```



```
  Query Match             0.3%;  Score 18;  DB 2;  Length 2524;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2486 AATGAAAAAAAAAAAAAA 2503


RESULT 174
US-08-728-626-1
; Sequence 1, Application US/08728626
; Patent No. 5910451
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,626
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/439,818
;      FILING DATE:  12-MAY-1995
;      APPLICATION NUMBER:  US 08/317,522
;      FILING DATE:  04-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1563
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2524 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```



```
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..2275
US-08-728-626-1


  Query Match             0.3%;  Score 18;  DB 2;  Length 2524;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2486 AATGAAAAAAAAAAAAAA 2503


RESULT 175
US-08-808-599A-1
; Sequence 1, Application US/08808599A
; Patent No. 6111089
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin, Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins and Methods to Inhibit Implantation
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/808,599A
;      FILING DATE:  28-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/317,522
;      FILING DATE:  04-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/439,818
;      FILING DATE:  12-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 2256
```



```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2524 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..2275
US-08-808-599A-1


  Query Match             0.3%;  Score 18;  DB 3;  Length 2524;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    2486 AATGAAAAAAAAAAAAAA 2503


RESULT 176
US-08-514-014-3
; Sequence 3, Application US/08514014
; Patent No. 5707829
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  Kelleher, Kerry
;    APPLICANT:  Carlin, McKeough
;    TITLE OF INVENTION:  DNA SEQUENCES AND SECRETED PROTEINS
;    TITLE OF INVENTION:  ENCODED THEREBY
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc. -- Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/514,014
;      FILING DATE:
```

```
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI6000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2582 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  52..2034
US-08-514-014-3


  Query Match             0.3%;  Score 18;  DB 1;  Length 2582;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   2563 AATGAAAAAAAAAAAAAA 2580


RESULT 177
US-08-833-823-3
; Sequence 3, Application US/08833823
; Patent No. 5969093
; GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  Kelleher, Kerry
;    APPLICANT:  Carlin, McKeough
;    TITLE OF INVENTION:  DNA SEQUENCES AND SECRETED PROTEINS
;    TITLE OF INVENTION:  ENCODED THEREBY
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc. -- Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
```



```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,823
;      FILING DATE:  10-APR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/514,014
;      FILING DATE:  11-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI6000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2582 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  52..2034
US-08-833-823-3


  Query Match             0.3%;  Score 18;  DB 2;  Length 2582;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   2563 AATGAAAAAAAAAAAAAA 2580


RESULT 178
US-08-178-477B-42
; Sequence 42, Application US/08178477B
; Patent No. 5756343
; GENERAL INFORMATION:
;    APPLICANT:  WU, CARL; CLOS, JOACHIM;
;    APPLICANT:  WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
;    TITLE OF INVENTION:  CELL STRESS
;    TITLE OF INVENTION:  TRANSCRIPTIONAL FACTORS
;    NUMBER OF SEQUENCES:  42
```



```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/178,477B
;      FILING DATE:  07-JAN-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/617,910
;      FILING DATE:  26-NOV-1990
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4103US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2781
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-178-477B-42


  Query Match             0.3%;  Score 18;  DB 1;  Length 2781;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   2753 AATGAAAAAAAAAAAAAA 2770


RESULT 179
US-08-747-221B-30
; Sequence 30, Application US/08747221B
; Patent No. 6063610
```



```
; GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6063610el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,221B
;      FILING DATE:  No. 6063610ember 12, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  PC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2801 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  99..1886
US-08-747-221B-30


  Query Match             0.3%;  Score 18;  DB 3;  Length 2801;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4558 aaagaaaaatcagaaaaa 4575
          ||||||||||||||||||
Db   1935 AAAGAAAAATCAGAAAAA 1952
```

```
RESULT 180
US-08-747-221B-32/c
; Sequence 32, Application US/08747221B
; Patent No. 6063610
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6063610el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,221B
;      FILING DATE:  No. 6063610ember 12, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2801 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-747-221B-32

  Query Match             0.3%;  Score 18;  DB 3;  Length 2801;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4558 aaagaaaaatcagaaaaa 4575
          ||||||||||||||||||
```



```
Db    867 AAAGAAAAATCAGAAAAA 850


RESULT 181
US-08-747-221B-24
; Sequence 24, Application US/08747221B
; Patent No. 6063610
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6063610el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,221B
;      FILING DATE:  No. 6063610ember 12, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2836 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  99..1889
US-08-747-221B-24

  Query Match             0.3%;  Score 18;  DB 3;  Length 2836;
```



```
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4558 aaagaaaaatcagaaaaa 4575
          ||||||||||||||||||
Db   1938 AAAGAAAAATCAGAAAAA 1955


RESULT 182
US-08-747-221B-26/c
; Sequence 26, Application US/08747221B
; Patent No. 6063610
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6063610el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,221B
;      FILING DATE:  No. 6063610ember 12, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2836 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-747-221B-26
```



```
  Query Match             0.3%;  Score 18;  DB 3;  Length 2836;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4558 aaagaaaaatcagaaaaa 4575
          ||||||||||||||||||
Db    899 AAAGAAAAATCAGAAAAA 882


RESULT 183
US-09-014-969-8
; Sequence 8, Application US/09014969
; Patent No. 5965397
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John M.
;    APPLICANT:  LaVallie, Edward R.
;    APPLICANT:  Racie, Lisa A.
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    APPLICANT:  Agostino, Michael J.
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/014,969
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sprunger, Suzanne A.
;      REGISTRATION NUMBER:  41,323
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8284
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2999 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-014-969-8


  Query Match              0.3%;  Score 18;  DB 2;  Length 2999;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   2981 AATGAAAAAAAAAAAAAA 2998


RESULT 184
US-08-920-234-1
; Sequence 1, Application US/08920234
; Patent No. 5922546
;  GENERAL INFORMATION:
;     APPLICANT:  ELLIS, CATHERINE
;     APPLICANT:  JACKSON, JEFFREY
;     APPLICANT:  MAYER, RUTH
;     TITLE OF INVENTION:  HUMAN DISINTEGRIN METALLOPROTEASE
;     TITLE OF INVENTION:  RELATED TO DROSOPHILA KUZ GENE
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  RATNER & PRESTIA
;       STREET:  P.O. BOX 980
;       CITY:  VALLEY FORGE
;       STATE:  PA
;       COUNTRY:  USA
;       ZIP:  19482
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/920,234
;       FILING DATE:  25-AUG-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PRESTIA, PAUL F
;       REGISTRATION NUMBER:  23,031
;       REFERENCE/DOCKET NUMBER:  GH-70237
;     TELECOMMUNICATION INFORMATION:
```



```
;       TELEPHONE:  610-407-0700
;       TELEFAX:  610-407-0701
;       TELEX:  846169
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3349 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-920-234-1


  Query Match              0.3%;  Score 18;  DB 2;  Length 3349;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   3279 AATGAAAAAAAAAAAAAA 3296


RESULT 185
US-08-529-654-3
; Sequence 3, Application US/08529654
; Patent No. 5739284
;  GENERAL INFORMATION:
;     APPLICANT:  HEDIGER, MATTHIAS
;     APPLICANT:  KANAI, YOSHIKATSU
;     TITLE OF INVENTION:  COMPOSITIONS CORRESPONDING TO A HIGH
;     TITLE OF INVENTION:  AFFINITY GLUTAMATE TRANSPORTER MOLECULE AND METHODS FOR
;     TITLE OF INVENTION:  MAKING AND USING SAME
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  WOLF, GREENFIELD & SACKS, P.C.
;       STREET:  600 ATLANTIC AVENUE
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  USA
;       ZIP:  02210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/529,654
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/194,719
```



```
;       FILING DATE:  10-FEB-1994
;       APPLICATION NUMBER:  US 07/965,676
;       FILING DATE:  19-OCT-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  JANIUK, ANTHONY J.
;       REGISTRATION NUMBER:  29,809
;       REFERENCE/DOCKET NUMBER:  B0801/7021
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-720-3500
;       TELEFAX:  617-720-2441
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3442 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  LEPORIDAE (RABBIT)
US-08-529-654-3


  Query Match              0.3%;  Score 18;  DB 1;  Length 3442;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    385 aatgaaaaaaaaaaaaaa 402
          ||||||||||||||||||
Db   3423 AATGAAAAAAAAAAAAAA 3440


RESULT 186
US-08-591-079-7/c
; Sequence 7, Application US/08591079
; Patent No. 5972899
;  GENERAL INFORMATION:
;     APPLICANT:  Zychlinsky, Arturo
;     APPLICANT:  Chen, Yajing
;     TITLE OF INVENTION:  Apoptosis Induced by Shigella IpaB
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORRISON & FOERSTER
;       STREET:  2000 Pennsylvania Avenue, NW
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20006-1812
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
```



```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/591,079
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Livnat, Shmuel
;       REGISTRATION NUMBER:  33,949
;       REFERENCE/DOCKET NUMBER:  15661-20017.00
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 887-1500
;       TELEFAX:  (202) 887-0764
;       TELEX:  90-4030-MRSNFOERSWSH
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3622 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     ORIGINAL SOURCE:
;       ORGANISM:  Salmonella typhimurium
;       STRAIN:  sibB
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  575..2356
;       OTHER INFORMATION:  /product= "SipB"
;       OTHER INFORMATION:  /gene= "sipB"
US-08-591-079-7


  Query Match              0.3%;  Score 18;  DB 2;  Length 3622;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4855 agaatatctttagcgcta 4872
          ||||||||||||||||||
Db   2490 AGAATATCTTTAGCGCTA 2473


RESULT 187
PCT-US96-00419-1
; Sequence 1, Application PC/TUS9600419
;  GENERAL INFORMATION:
;     APPLICANT:  Thomas Ciossek, Axel Ullrich, Birgit
;     APPLICANT:  Millauer
;     TITLE OF INVENTION:  METHODS FOR DIAGNOSIS AND TREATMENT OF
;     TITLE OF INVENTION:  MDK1 SIGNAL TRANSDUCTION DISORDERS
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Lyon & Lyon
;       STREET:  633 West Fifth Street
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM compatible
;       OPERATING SYSTEM:  IBM P.C. DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US96/00419
;       FILING DATE:  January 3, 1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;     PRIOR APPLICATION DATA:  including application
;     PRIOR APPLICATION DATA:  described below:                    none
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  208/007
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4304
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  nucleic
PCT-US96-00419-1


  Query Match           0.3%;  Score 18;  DB 4;  Length 4304;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    3981 AATGAAAAAAAAAAAAAA 3998


RESULT 188
US-08-591-079-9/c
; Sequence 9, Application US/08591079
; Patent No. 5972899
```



```
;  GENERAL INFORMATION:
;    APPLICANT:  Zychlinsky, Arturo
;    APPLICANT:  Chen, Yajing
;    TITLE OF INVENTION:  Apoptosis Induced by Shigella IpaB
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,079
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Livnat, Shmuel
;      REGISTRATION NUMBER:  33,949
;      REFERENCE/DOCKET NUMBER:  15661-20017.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0764
;      TELEX:  90-4030 MRSNFOERSWSH
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Salmonella typhi
;      STRAIN:  Ty2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  543..2324
;      OTHER INFORMATION:  /gene= "sipB"
US-08-591-079-9.


  Query Match           0.3%;  Score 18;  DB 2;  Length 5393;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    4855 agaatatctttagcgcta 4872
```



```
           ||||||||||||||||||
Db    2458 AGAATATCTTTAGCGCTA 2441


RESULT 189
US-08-757-439-1/c
; Sequence 1, Application US/08757439
; Patent No. 5866371
;  GENERAL INFORMATION:
;    APPLICANT:  BADIONG, Werner
;    APPLICANT:  HABERMANN, Paul
;    APPLICANT:  MOELLER, Joerg
;    APPLICANT:  ARETZ, Werner
;    TITLE OF INVENTION:  PROCESS FOR USING THE YEAST ADH II
;    TITLE OF INVENTION:  PROMOTER SYSTEM FOR THE PRODUCTION OF HETEROLOGOUS
;    TITLE OF INVENTION:  PROTEINS IN HIGH YIELDS
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/757,439
;      FILING DATE:  27-NOV-1996
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19544233.4
;      FILING DATE:  28-NOV-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SANDERCOCK, Colin G.
;      REGISTRATION NUMBER:  31,298
;      REFERENCE/DOCKET NUMBER:  18748/303/HOCE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8491 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-757-439-1
```



```
  Query Match           0.3%;  Score 18;  DB 2;  Length 8491;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     385 aatgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db    7502 AATGAAAAAAAAAAAAAA 7485


RESULT 190
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tauchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237e1
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-20


  Query Match           0.3%;  Score 18;  DB 2;  Length 246240;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgct 406
          ||||||||||||||||||
Db  63123 AAAAAAAAAAAAAATGCT 63140


RESULT 191
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237e1
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
```



```
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-21


  Query Match           0.3%;  Score 18;  DB 2;  Length 246240;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgct 406
          ||||||||||||||||||
Db  63123 AAAAAAAAAAAAAATGCT 63140


RESULT 192
US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237e1
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
```



```
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-22


  Query Match           0.3%;  Score 18;  DB 2;  Length 246240;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    389 aaaaaaaaaaaaaatgct 406
          ||||||||||||||||||
Db  63123 AAAAAAAAAAAAAATGCT 63140


RESULT 193
US-08-553-367A-17
; Sequence 17, Application US/08553367A
; Patent No. 5939539
;  GENERAL INFORMATION:
;    APPLICANT:  Theodor LANGE et al.
;    TITLE OF INVENTION:  GA 20-OXIDASE GENE SEQUENCES (As Amended)
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack, L.L.P.
```



```
;      STREET:  2033 K Street, N.W., Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/553,367A
;      FILING DATE:  No. 5939539ember 27, 1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee Cheng
;      REGISTRATION NUMBER:  40,949
;      REFERENCE/DOCKET NUMBER:  49/FD4.5MZ
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  Oligo 3  (Beispiel 39b)
US-08-553-367A-17


  Query Match           0.3%;  Score 17;  DB 2;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 5e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2641 gaattcaaaaatggcaa 2657
          |||||||||||||||||
Db      3 GAATTCAAAAATGGCAA 19


RESULT 194
US-08-132-990A-25/c
; Sequence 25, Application US/08132990A
```

; Patent No. 5834589
; GENERAL INFORMATION:
; APPLICANT: MERUELO, DANIEL
; APPLICANT: YOSHIMOTO, TAKAYUKI
; TITLE OF INVENTION: Human Retrovirus Receptor and DNA Coding Therefor
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/132,990A
; FILING DATE: 07-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/084,729
; FILING DATE: 29-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05569
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/899,075
; FILING DATE: 11-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/806,178
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/627,950
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 8105-004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA

US-08-132-990A-25

Query Match 0.3%; Score 17; DB 2; Length 32;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5777 taaaccctcccacaccg 5793
||||||||||||||||| .
Db 31 TAAACCCTCCCACACCG 15

RESULT 195
US-08-961-083-330/c
; Sequence 330, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Andera
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 330:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

; TOPOLOGY: linear
US-08-961-083-330

Query Match 0.3%; Score 17; DB 3; Length 33;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2238 agaagctaataagctta 2254
||||||||||||||||| .
Db 20 AGAAGCTAATAAGCTTA 4

RESULT 196
US-08-418-848A-57/c
; Sequence 57, Application US/08418848A
; Patent No. 5847096
; GENERAL INFORMATION:
; APPLICANT: SCHUBERT, MANFRED, HARMISON II,
; APPLICANT: GEORGE G., CHANG-JIE, CHEN, BANJERJEA, AKHIL
; TITLE OF INVENTION: DEFECTIVE, INTERFERING
; TITLE OF INVENTION: HIV PARTICLES
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,848A
; FILING DATE: 07-APR-1995
; CLASSIFICATION: 526
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/936,849
; FILING DATE: 28-AUG-1992
; CLASSIFICATION: 526
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4091US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 57:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-418-848A-57

Query Match 0.3%; Score 17; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 389 aaaaaaaaaaaaaatgc 405
||||||||||||||||| .
Db 27 AAAAAAAAAAAAAATGC 11

RESULT 197
US-08-120-827-97
; Sequence 97, Application US/08120827
; Patent No. 5525495
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,827
; FILING DATE: 15-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5525495man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:

```
;      TELEPHONE:  (703)413-3000
;      TELEFAX:  (703)413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  97:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  85 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-120-827-97


  Query Match             0.3%;  Score 17;  DB 1;  Length 85;
  Best Local Similarity  100.0%;  Pred. No. 4.6e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     386 atgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      48 ATGAAAAAAAAAAAAAA 64


RESULT 198
US-08-478-675-97
; Sequence 97, Application US/08478675
; Patent No. 5773246
;  GENERAL INFORMATION:
;    APPLICANT:  KEENE, JACK D.
;    APPLICANT:  KING, PETER H.
;    APPLICANT:  LEVINE, TODD
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;    TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
;    TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,675
;      FILING DATE:  07-JUN-1996
;      CLASSIFICATION:  536
```



```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/120,827
;      FILING DATE:  15-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5773246man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)413-3000
;      TELEFAX:  (703)413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  97:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  85 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-478-675-97


  Query Match             0.3%;  Score 17;  DB 1;  Length 85;
  Best Local Similarity  100.0%;  Pred. No. 4.6e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     386 atgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      48 ATGAAAAAAAAAAAAAA 64


RESULT 199
US-08-735-381-2
; Sequence 2, Application US/08735381
; Patent No. 5853993
;  GENERAL INFORMATION:
;    APPLICANT:  Dellinger, Douglas J.
;    APPLICANT:  Dahm, SueAnn
;    APPLICANT:  Troll, Mark
;    TITLE OF INVENTION:  SIGNAL ENHANCEMENT METHOD AND KIT
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hewlett-Packard Company, Legal Dept.,
;      ADDRESSEE:  Intellectual Property
;      STREET:  1501 Page Mill Road, MS 4U-10
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1126
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```



```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/735,381
;      FILING DATE:  21-OCT-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Choi, Wendy A.
;      REGISTRATION NUMBER:  36,697
;      REFERENCE/DOCKET NUMBER:  10950427-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-857-4125
;      TELEFAX:  650-852-8063
;      TELEX:  348-461
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  105 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-735-381-2


  Query Match             0.3%;  Score 17;  DB 2;  Length 105;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     386 atgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      53 ATGAAAAAAAAAAAAAA 69


RESULT 200
US-09-183-619-1
; Sequence 1, Application US/09183619
; Patent No. 6103474
; GENERAL INFORMATION:
;  APPLICANT: DELLINGER, DOUGLAS J.
;  APPLICANT: DAHM, SUEANN C.
;  APPLICANT: ILSLEY, DIANE D.
;  APPLICANT: ACH, ROBERT A.
;  APPLICANT: TROLL, MARK A.
;  TITLE OF INVENTION: HYBRIDIZATION ASSAY SIGNAL ENHANCEMENT
;  FILE REFERENCE: 10981619-1
;  CURRENT APPLICATION NUMBER: US/09/183,619
;  CURRENT FILING DATE: 1998-10-30
;  EARLIER APPLICATION NUMBER: 08/735,381
;  EARLIER FILING DATE: 1996-10-21
```



```
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 105
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Target analyte
US-09-183-619-1


  Query Match             0.3%;  Score 17;  DB 3;  Length 105;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     386 atgaaaaaaaaaaaaaa 402
           ||||||||||||||||||
Db      53 atgaaaaaaaaaaaaaa 69


Search completed: March 26, 2001, 23:07:10
Job time: 31617 sec
```